

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

Om protein - protein search, using sw model

Run on: July 30, 2002, 14:47:39 ; Search time 53.42 Seconds  
(without alignments)  
95.646 Million cell updates/sec

Title: US-09-833-017-2  
Perfect score: 227  
Sequence: 1 MRRKLSKNDKEIKDELE.....GSLSTPFRLNRSTQALGK 46

Scoring table: BLOSUM62  
Gapext 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing:  
Minimum Match 0 %  
Listing first 45 summaries

Database : A\_GeneSeq\_032802:\*

1: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1980.DAT; \*  
2: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1981.DAT; \*  
3: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1982.DAT; \*  
4: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1983.DAT; \*  
5: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1984.DAT; \*  
6: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1985.DAT; \*  
7: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1987.DAT; \*  
8: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1988.DAT; \*  
9: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1989.DAT; \*  
10: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1990.DAT; \*  
11: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1992.DAT; \*  
12: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1991.DAT; \*  
13: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1993.DAT; \*  
14: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1994.DAT; \*  
15: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1995.DAT; \*  
16: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1996.DAT; \*  
17: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1997.DAT; \*  
18: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1998.DAT; \*  
19: /SIDS1/godata/hold-geneseq/geneseq-embl/AA1999.DAT; \*  
20: /SIDS1/godata/hold-geneseq/geneseq-embl/AA2000.DAT; \*  
21: /SIDS1/godata/hold-geneseq/geneseq-embl/AA2001.DAT; \*  
22: /SIDS1/godata/hold-geneseq/geneseq-embl/AA2002.DAT; \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			ALIGNMENTS		
Result No.	Score	Match Length	DB ID	Description	
1	58	25.6	173	18 RAW28307	<i>Staphylococcus aureus</i>
2	56	24.7	939	22 AAU35204	<i>Enterococcus faecalis</i>
3	55.5	24.4	553	8 AAB70176	Sequence of Newcastle disease
4	55.5	24.4	553	15 AAB59858	Newcastle disease
5	55.5	24.4	553	15 AAB49341	Newcastle disease
6	55.5	24.4	553	17 AAW06228	Newcastle disease
7	55.5	24.4	553	18 AAW10691	Newcastle disease
8	55.5	24.4	553	20 AAY21983	Sequencing ID No: 14 of U Protein encoded by
9	55.5	24.4	553	21 AAB36039	Newcastle disease
10	55.5	24.4	553	21 AAY51231	NDV fusion (F) protein
11	55.5	24.4	553	21 AAY58183	

RESULT 1  
ID AAW28307 standard; Protein; 173 AA.  
AC AAW28307;  
XX  
DT 01-SEP-1998 (first entry)  
DE *Staphylococcus aureus* protein of unknown function.  
XX  
KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;  
KW Staphylococcal gene; regulatory element; bacterial gene expression;  
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;  
KW toxic shock syndrome.  
OS Staphylococcus aureus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 123  
FT Misc-difference 132  
FT Misc-difference /note= "X is not defined in the specification"  
FT Misc-difference /note= "X is not defined in the specification"  
XX  
PN W09730070-A1.  
XX  
PD 21-AUG-1997.  
XX  
PF 19-FEB-1997; 97WO-US023118.  
XX  
PR 20-FEB-1996; 96US-0011888.  
XX  
PA (SMIC) SMITHKLINE BEECHAM CORP.  
XX  
PI Black MT, Burnham MK, Hodgson JE, Knowles DIC, Nicholas RO;

S. epidermidis ope  
Arabidopsis thalia  
Arabidopsis thaliana  
Monellin protein.  
Haemophilus adhesi  
Haemophilus influenzae  
Newcastle disease  
Arabidopsis thalia  
L. lactis branched  
Novel human diagno  
Human protein SEQ  
Human protein SEQ  
Human protein SEQ  
M. jannaschii MJ08  
Novel human diagno  
Novel human diagno  
Bacillus cereus  
Escherichia coli P  
E. coli proliferat  
E. coli growth and  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Staphylococcus aur  
Human colon cancer  
Mutant uracil DNA  
Mutant uracil DNA  
Mutant uracil DNA  
Mutant uracil DNA  
Uracil DNA glycosy  
Cytosine DNA glyco  
Thymine DNA glyco  
Mutant uracil DNA

PT Pratt JM, Reichard RW, Rosenberg M, Ward JM;  
 XX WPI; 1997-424959/39.  
 DR N-PSDB; AAT84204.

PT Novel polypeptide(s) from *Staphylococcus aureus* strain WCUH29 - used  
 XX to isolate antimicrobial compounds, and in vaccines against *S. aureus* infection  
 PR  
 PS Claim 6; Page 597-598; 989pp; English.

CC The present sequence represents a *Staphylococcus aureus* protein of  
 CC unknown function. The DNA sequence was isolated from a library of  
 CC clones of *S. aureus* WCUH 29 in *Escherichia coli*. The DNA sequence can  
 CC be used in the construction of ribozymes and antisense sequences to  
 CC control the expression of *Staphylococcal* genes. The DNA sequence is  
 CC also useful as a source of regulatory elements for the control of  
 CC bacterial gene expression. The present protein may be used to produce  
 CC vaccines to enable a host to produce specific antibodies with  
 CC antibacterial action. These vaccines and antibiotics would protect  
 CC a host against invasion by *S. aureus*, and conditions relating to  
 CC *Staphylococcal* infection, e.g. *Staphylococcal* food poisoning, scaled  
 CC skin syndrome, and toxic shock syndrome.

SQ Sequence 173 AA;

Query Match 25.6%; Score 58; DB 18; Length 173;  
 Best Local Similarity 34.3%; Pred. No. 3.3; Matches 8; Mismatches 15; Indels 0; Gaps 0;  
 AC AAU35204;  
 XX AAU35204;  
 DT 13-FEB-2002 (first entry)  
 DE Enterococcus faecalis cellular proliferation protein #491.  
 XX  
 KW Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Enterococcus faecalis.  
 XX  
 WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PP 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242278P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257331P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELTRA PHARM INC.  
 XX  
 PI Hasebeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI; 2001-611495/70.  
 DR N-PSDB; AAS53063.

PT New polynucleotides for the identification and development of  
 XX antibiotics, comprise sequences of antisense nucleic acids .  
 XX  
 PS Example 3; Seq ID No 10797; 511PB; English.

CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification for this patent, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp://ipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 939 AA;

Query Match 24.7%; Score 56; DB 22; Length 939;  
 Best Local Similarity 33.3%; Pred. No. 45; Matches 11; Conservative 7; Mismatches 15; Indels 0; Gaps 0;  
 AC AAP70176;  
 XX AAP70176 standard; protein; 553 AA.  
 DT 03-APR-1991 (first entry)  
 DE Sequence of Newcastle Disease Virus (NDV) F polypeptide.  
 XX  
 KW Epitope; probe; diagnosis.  
 XX  
 OS Newcastle Disease Virus.  
 XX  
 PN EP227414-A.  
 XX  
 PD 01-JUL-1987.  
 XX  
 PP 16-DEC-1986; 86EP-0309804.  
 XX  
 PR 15-JUL-1986; 86US-0885765.  
 PR 18-DEC-1985; 85GB-0031147.  
 PR 14-APR-1986; 86GB-0009037.  
 XX  
 PA (NABR ) NATIONAL RES DEV CORP.  
 XX  
 PI Bingham RW, Chambers P, Emerson PR, Millar NS;  
 XX  
 DR WPI; 1987-179630/26.  
 XX  
 PS Example; pages 11-16; 22pp; English.

PT Newcastle disease virus gene clones - comprise polynucleotide(s)  
 XX encoding the HN and/or F protein of Newcastle disease virus RNA  
 PT  
 DR N-PSDB; AAN70661.



XX  
XX DT 04-MAR-1997 (first entry)  
XX DE Newcastle disease virus F gene product.  
XX KW Turkey herpes virus; recombinant virus; vaccine; prophylaxis;  
KW immunisation; avian virus; infectious bronchitis virus;  
KW infectious bursal disease virus; Newcastle disease virus;  
KW Marek's disease virus; infectious laryngotracheitis virus; IBV;  
KW IBDV; NDV; MDV; ILV.  
XX OS Newcastle disease virus.  
XX WO9605291-A1.  
XX PD 22-FEB-1996.  
XX PF 09-AUG-1995; 95WO-US10245.  
XX PR 22-DEC-1994; 94US-0362240.  
XX PR 09-AUG-1994; 94US-0288065.  
XX PA (SYTR ) SYNTRO CORP.  
XX PI Cochran MD, Junker DE, Singer PA;  
XX DR WPI: 1997-087060/0B.  
XX N-PSDB; AAT18203.  
XX Disclosure; Page 179-180; 249pp; English.  
XX PT Recombinant turkey herpes viruses contg. foreign DNA encoding a  
PT cytokine - useful in vaccines to protect against Marek's disease  
PT virus and other avian viruses.  
XX PS Disclosure; Page 109-110; 134pp; English.  
XX CC Recombinant turkey herpes virus (rTH) which comprise a foreign DNA  
CC sequence encoding a cytokine inserted into a XbaI site within an  
ECORI #9 genomic fragment, where the cytokine can be expressed in  
CC host cells infected with the virus can be used in vaccines to  
protect turkeys against avian viruses. The recombinant viruses can  
be used for immunising birds against infectious bronchitis virus  
(IBV), infectious laryngotracheitis virus (ILV), Marek's disease virus  
(MDV), infectious bursal disease virus (IBDV), and Newcastle disease  
virus (NDV). They may also be used in multivalent vaccines to  
protect against one or more of these avian viruses. This sequence  
is the product of the F gene of the Newcastle disease virus and is  
an antigen which can be used in the recombinant vaccines.  
XX Sequence . 553 AA;  
SQ Query Match 24.4%; Score 55.5; DB 17; Length 553;  
Best Local Similarity 33.3%; Pred. No. 29; Mismatches 7; Indels 5; Gaps 1;  
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;  
Qy OY 4 TLSLKNDF----KEKTDPELEIIGGSGSLSLTFRLNRSFTQALGK 46  
Db 433 tirlsgefdrvtyqkniqdsqvlgtgnldistelgnvnnslnlk 480  
RESULT 8  
ID AAY21983  
ID AAY21983 standard; Protein; 553 AA.  
XX AC AAY21983;  
XX DT 07-SEP-1999 (first entry)  
XX DE Seq ID No: 14 of US5925358.  
XX KW Fowlpox virus; FPV; recombinant; vaccine; immunisation; chicken; NDV;  
KW Newcastle disease virus; NDV; Fowlpox; Infectious Laryngotracheitis.  
XX OS Fowlpox virus.  
XX PN US5925358-A.  
XX PD 20-JUL-1999.  
XX PR 07-JUN-1995; 95US-0484575.  
XX PR 07-JUN-1995; 95US-0484575.  
PR 26-FEB-1993; 93US-0024156.  
PR 28-FEB-1994; 94WO-US02252.  
XX PA (SYTR ) SYNTRO CORP.  
XX PI Cochran MD, Junker DE;

XX OS Newcastle disease virus.  
XX PN WO9640880-A1.  
XX PD 19-DEC-1996.  
XX PR 04-JUN-1996; 96WO-US11187.  
XX PR 07-JUN-1995; 95US-0484790.  
XX PA (SYTR ) SYNTRO CORP.  
XX PI Cochran MD, Junker DE, Singer PA;  
XX DR WPI: 1997-087060/0B.  
XX N-PSDB; AAT18203.  
XX Disclosure; Page 109-110; 134pp; English.  
XX CC Newcastle disease virus (NDV) haemagglutinin (HN) (AAW10690) and  
CC fusion (F) protein (AAW10691) are expressed by novel recombinant  
CC fowlpox virus (FPV). The genes (see also AAT48510) for HN and F can  
CC be inserted into homology vector 443-88.8 (see also AAT46511) at the  
CC unique SfiI site, yielding homology vector 502-26.22 (see also  
CC AAT45502-05), which can be used to insert the NDV HN and F genes  
CC into fowlpox virus. The recombinant virus is used to deliver the  
CC vaccine antigens to poultry.  
XX Sequence 553 AA;

Query Match 24.4%; Score 55.5; DB 18; Length 553;  
Best Local Similarity 33.3%; Pred. No. 29; Mismatches 7; Indels 5; Gaps 1;  
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;  
Qy OY 4 TLSLKNDF----KEKTDPELEIIGGSGSLSLTFRLNRSFTQALGK 46  
Db 433 tirlsgefdrvtyqkniqdsqvlgtgnldistelgnvnnslnlk 480







Query	Match	Score	DB	Length
Best Local	Similarity	24.0%	54.5%	508
Matches 21;	Conservative	39.6%	Pred. No. 36;	
			Mismatches 8;	
			Indels 11;	
			Gaps 13;	

PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139889.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151458.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0152367.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	13-SEP-1999;	99US-0154019.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154309.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142930.	PR	24-SEP-1999;	99US-0155307.
PR	08-JUL-1999;	99US-0142820.	PR	24-SEP-1999;	99US-0155659.
PR	12-JUL-1999;	99US-0142877.	PR	28-SEP-1999;	99US-0156458.
PR	13-JUL-1999;	99US-0143277.	PR	05-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-014324.	PR	05-OCT-1999;	99US-0157551.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144885.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144886.	PR	08-OCT-1999;	99US-0158237.
PR	19-JUL-1999;	99US-0144225.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144334.	PR	18-OCT-1999;	99US-0159765.
PR	19-JUL-1999;	99US-0144335.	PR	21-OCT-1999;	99US-0159339.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159350.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	13-OCT-1999;	99US-0159637.
PR	21-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160814.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160854.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160980.
PR	26-JUL-1999;	99US-0145276.	PR	22-OCT-1999;	99US-0160989.
PR	27-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-016104.
PR	27-JUL-1999;	99US-0145389.	PR	25-OCT-1999;	99US-016104.
PR	27-JUL-1999;	99US-0145389.	PR	25-OCT-1999;	99US-016104.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-016104.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-016104.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161085.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161920.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147620.			
PR	06-AUG-1999;	99US-01477303.			
PR	06-AUG-1999;	99US-01477303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	10-AUG-1999;	99US-0147935.			
PR	11-AUG-1999;	99US-0148171.			
PR	12-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	12-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149375.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			

PR 29-OCT-1999: 9911S-01631A3

XX  
 DE Monellin protein.  
 XX  
 KW Sweet; heat; resistance; protease.  
 XX  
 OS Synthetic.  
 XX  
 FH Location/Qualifiers  
 FT Misc-difference 51  
 FT /label= GLY, SER, GLU  
 XX  
 JP05070494-A.  
 XX  
 PD 23-MAR-1993.  
 XX  
 PF 30-MAY-1991; 91JP-0155713.  
 XX  
 PR 25-JUL-1990; 90JP-0196983.  
 XX  
 PA (KIRI ) KIRIN BREWERY KK.  
 XX  
 DR WPI; 1993-131302/16.  
 XX  
 PT Single-stranded monellin protein - has high heat resistance in  
 PT cow pH range and high protease resistance  
 XX  
 PS Claim 1; Page 2; 29pp; Japanese.  
 XX  
 CC Monellin, is a sweet protein, having high thermal resistance at low  
 CC pH's and also a high protease resistance.  
 XX  
 SQ Sequence 96 AA;

Query Match 23.8%; Score 54; DB 14; Length 96;  
 Best Local Similarity 41.4%; Pred. No. 6;  
 Matches 12; Conservative 6; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MKKTLSLKNDFFERIKTPELEIIGGSSL 29  
 ||||: :| |:| |:| :| :| |:  
 Db 42 mkktyeenxfreikgyeqlyyyasdkl 70

Search completed: July 30, 2002, 15:14:11  
 Job time: 1592 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model.

Run on: July 30, 2002, 15:06:04 ; Search time 23.35 Seconds  
(without alignments)

Title: US-09-833-017-2  
Perf. score: 227  
Sequence: 1 MKKTKSLKNDKFKEKTDELE.....GSLSITFFRLNRSPTQNLGK 46

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%, Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn1\_6/podata/2/1aa/5A\_COMB.pep:\*
- 2: /cgn1\_6/podata/2/1aa/5B\_COMB.pep:\*
- 3: /cgn1\_6/podata/2/1aa/6A\_COMB.pep:\*
- 4: /cgn1\_6/podata/2/1aa/6B\_COMB.pep:\*
- 5: /cgn1\_6/podata/2/1aa/PCUS.Comb.pep:\*
- 6: /cgn1\_6/podata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	55.5	24.4	553	2 US-09-663-566A-13 Sequence 13, Appl
2	55.5	24.4	553	2 US-09-484-575A-14 Sequence 14, Appl
3	55.5	24.4	553	2 US-09-023-610-13 Sequence 13, Appl
4	55.5	24.4	553	2 US-09-065A-13 Sequence 13, Appl
5	55.5	24.4	553	2 US-09-362-240A-13 Sequence 13, Appl
6	55.5	24.4	553	3 US-08-477-459-14 Sequence 14, Appl
7	55.5	24.4	553	3 US-08-479-869-14 Sequence 14, Appl
8	55.5	24.4	553	4 US-09-486-414-14 Sequence 14, Appl
9	55.5	24.4	553	4 US-09-804-372A-11 Sequence 11, Appl
10	55.5	24.4	553	5 PCT-US94-022526-14 Sequence 14, Appl
11	55.5	24.4	553	5 PCT-US95-10245-13 Sequence 13, Appl
12	55.5	24.4	553	6 5310678-1 Patent No. 5310678
13	55.5	24.4	837	4 US-09-564-805-228 Sequence 228, Appl
14	55.5	24.4	97	4 US-09-242-690A-6 Sequence 6, Appl
15	54	23.8	1912	1 US-08-409-995-4 Sequence 1, Appl
16	54	23.8	1912	3 US-08-685-467-4 Sequence 14, Appl
17	54	23.8	2353	4 US-09-669-974-33 Sequence 13, Appl
18	54	23.8	2353	4 US-08-919-942-4 Sequence 4, Appl
19	54	23.8	2353	4 US-08-919-942-4 Sequence 4, Appl
20	54	23.8	2353	4 US-09-669-974-33 Sequence 4, Appl
21	54	23.8	2353	4 US-09-669-974-33 Sequence 4, Appl
22	54	23.8	2411	4 US-09-268-347-36 Sequence 36, Appl
23	53	23.3	513	1 US-08-403-866-1 Sequence 1, Appl
24	51.5	22.7	244	4 US-08-919-573-2 Sequence 2, Appl
25	51.5	22.7	244	4 US-08-919-573-4 Sequence 4, Appl
26	49.5	21.8	436	3 US-08-466-099-94 Sequence 94, Appl
27	49.5	21.8	436	3 US-08-360-107A-104 Sequence 104, Appl

## ALIGNMENTS

RESULT 1	US-08-663-566A-13	SEQUENCE 94, Appl
	; Sequence 13, Application US/08663566A	Sequence 94, Appl
	; Patent No. 5853733	Sequence 94, Appl
	; GENERAL INFORMATION:	Sequence 94, Appl
	; APPLICANT: Cochran, Mark D	Sequence 94, Appl
	; TITLE OF INVENTION: Recombinant Hepesvirus of Turkeys	Sequence 94, Appl
	; TITLE OF INVENTION: and Uses Thereof	Sequence 94, Appl
	; NUMBER OF SEQUENCES: 56	Sequence 94, Appl
	; CORRESPONDENCE ADDRESS:	Sequence 94, Appl
	; ADDRESSEE: John P. White	Sequence 94, Appl
	; STREET: 1185 Avenue of the Americas	Sequence 94, Appl
	; CITY: New York	Sequence 94, Appl
	; STATE: New York	Sequence 94, Appl
	; COUNTRY: USA	Sequence 94, Appl
	; ZIP: 10036	Sequence 94, Appl
	; COMPUTER READABLE FORM:	Sequence 94, Appl
	; COMPUTER: IBM PC compatible	Sequence 94, Appl
	; OPERATING SYSTEM: MS-DOS	Sequence 94, Appl
	; SOFTWARE: PatentIn Release #1.0, Version #1.25	Sequence 94, Appl
	; CURRENT APPLICATION DATA:	Sequence 94, Appl
	; APPLICATION NUMBER: US/08663,566A	Sequence 94, Appl
	; FILING DATE: June 13, 1996	Sequence 94, Appl
	; CLASSIFICATION: 435	Sequence 94, Appl
	; ATTORNEY/AGENT INFORMATION:	Sequence 94, Appl
	; NAME: White, John P	Sequence 94, Appl
	; REGISTRATION NUMBER: 28,678	Sequence 94, Appl
	; TELECOMMUNICATION INFORMATION:	Sequence 94, Appl
	; TELEPHONE: (212)278-0400	Sequence 94, Appl
	; TELEFAX: (212)391-0526	Sequence 94, Appl
	; TELEX: 422523	Sequence 94, Appl
	; INFORMATION FOR SEQ ID NO: 13:	Sequence 94, Appl
	; INFORMATION FOR SEQ ID NO: 13:	Sequence 94, Appl
	; SEQUENCE CHARACTERISTICS:	Sequence 94, Appl
	; LENGTH: 553 amino acids	Sequence 94, Appl
	; TYPE: amino acid	Sequence 94, Appl
	; TOPOLOGY: linear	Sequence 94, Appl
	; MOLECULE TYPE: protein	Sequence 94, Appl
	; US-08-663-566A-13	Sequence 94, Appl
	Query Match Best Local Similarity 24.4%; Score 55.5; DB 2; Length 553;	Sequence 94, Appl
	Matches 16; Conservative 33.3%; Pred. No. 3.4; Mismatches 20; Indels 5; Gaps 1;	Sequence 94, Appl
QY	4 TLSKNDFE---KEIKTDELEITIGGSGSLSTFFRLNRSPTQNLGK 46	Sequence 94, Appl
Db	4 33 TLRSGEFDVYKNIQSIODSQVITGNLDISTELGVNNNSNALNK 480	Sequence 94, Appl



RESULT 5  
US-08-362-240A-13  
Sequence 13, Application US/08362240A  
; Patent No. 5905138

GENERAL INFORMATION:  
APPLICANT: Cochran, Mark D  
APPLICANT: Juniper, David  
APPLICANT: Wild, Martha A  
TITLE OF INVENTION: Recombinant Herpesvirus and Uses Thereof  
NUMBER OF SEQUENCES: 60

CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/362,240A  
FILING DATE: Dec-22-94  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)278-0400  
TELEFAX: (212)391-0526  
TELEX: 422523

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:  
LENGTH: 553 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-362-240A-13

---

RESULT 6  
US-08-477-459-14  
Sequence 14, Application US/08477459  
; Sequence 14, Application US/08477459  
; Patent No. 6001369

GENERAL INFORMATION:  
APPLICANT: Mark D. Cochran  
TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses  
NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,869  
FILING DATE:  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/024,156  
FILING DATE: 26-FEB-1993

ATTORNEY/AGENT INFORMATION:  
NAME: White, Esq., John P.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)977-9550  
TELEFAX: (212)664-0525  
TELEX: 422523

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
LENGTH: 553 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-477-459-14

---

Query Match 24.4%; Score 55.5; DB 2; Length 553;  
Best Local Similarity 33.3%; Pred. No. 3.4; Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

QY 4 TISLKNDF----KEIKDDELTIGGGSLSFFRERNSFTQALK 46  
Db 433 TRLSGEFDVTVQKNISIQDSQVITGNLDISTELGNVNNSISNALNK 480

---

RESULT 7  
US-08-479-869-14  
Sequence 14, Application US/08479869  
; Sequence 14, Application US/08479869  
; Patent No. 6123949

GENERAL INFORMATION:  
APPLICANT: Cochran, Ph.D., Mark D  
TITLE OF INVENTION: Recombinant Fowlpox Virus S-FPV-043 and  
TITLE OF INVENTION: Uses Thereof  
NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,869  
FILING DATE:  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/024,156  
FILING DATE: 26-FEB-1993

ATTORNEY/AGENT INFORMATION:  
NAME: White, Esq., John P.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)977-9550  
TELEFAX: (212)664-0525  
TELEX: 422523

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
LENGTH: 553 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-479-869-14

---

Query Match 24.4%; Score 55.5; DB 3; Length 553;  
Best Local Similarity 33.3%; Pred. No. 3.4;

Matches 16; conservative 7; Mismatches 20; Indels 5; Caps 1;

RESULT 8 ; TYPE: amino acid  
 QY 4 TLSIKNDF----KEIKTDELEIITGGSGSLSTFRLFNRSFTQALGK 46 ; TOPOLOGY: linear  
 US-08-486-414-14 ; MOLECULE TYPE: protein  
 ; Sequence 14, Application US/08486414B  
 ; PATENT NO. 6136318  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cochran, Mark D.  
 ; ATTORNEY: Junker, David E.  
 ; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUSES AND USES THEREOF  
 ; FILE REFERENCE: 4271D  
 ; CURRENT APPLICATION NUMBER: US/08/486,414B  
 ; NUMBER OF SEQ ID NOS: 46  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO: 14  
 ; LENGTH: 553  
 ; TYPE: PRT  
 ; ORGANISM: Fowlpox virus  
 ; US-08-486-414-14

Query Match 24.4%; Score 55.5; DB 4; Length 553;  
 Best Local Similarity 33.3%; Pred. No. 3.4;  
 Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

RESULT 9 ; TYPE: amino acid  
 US-08-804-372A-11 ; TOPOLOGY: linear  
 ; Sequence 11, Application US/08804372A  
 ; PATENT NO. 6183753  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cochran, Mark D.  
 ; ATTORNEY: Wild, Martha A.  
 ; TITLE OF INVENTION: Recombinant Chimeric Viruses and Uses Thereof  
 ; NUMBER OF SEQUENCES: 32  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham LLP  
 ; STREET: 1185 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10112  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US94/01826A  
 ; FILING DATE: 28-FEB-1994  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White Esq, John P  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212)977-9550  
 ; TELEFAX: (212)664-0525  
 ; TELEX: 422523  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 553 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; PCT-US94-01826A-14

Query Match 24.4%; Score 55.5; DB 5; Length 553;  
 Best Local Similarity 33.3%; Pred. No. 3.4;  
 Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

RESULT 10 ; TYPE: amino acid  
 PCT-US94-01826A-14 ; TOPOLOGY: linear  
 ; Sequence 14, Application PCT/US9401826A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Syntro Corporation, et al.  
 ; TITLE OF INVENTION: Recombinant Fowlpox Virus S-FPV-043 and Uses Thereof  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: John P. White  
 ; STREET: 30 Rockefeller Plaza  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10112  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US94/01826A  
 ; FILING DATE: 28-FEB-1994  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White Esq, John P  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212)977-9550  
 ; TELEFAX: (212)664-0525  
 ; TELEX: 422523  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 553 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; PCT-US94-01826A-14

Query Match 24.4%; Score 55.5; DB 5; Length 553;  
 Best Local Similarity 33.3%; Pred. No. 3.4;  
 Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

RESULT 11 ; TYPE: amino acid  
 PCT-US94-02252A-14 ; TOPOLOGY: linear  
 ; Sequence 14, Application PCT/US9402252A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Syntro Corporation, et al.  
 ; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses Thereof  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: John P. White  
 ; STREET: 30 Rockefeller Plaza  
 ; CITY: New York

STATE: New York  
 COUNTRY: USA  
 ZIP: 10112  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT-US94/02252A  
 FILING DATE: 28-FEB-1994  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: White Esa, John P  
 TELECOMMUNICATION INFORMATION:  
 TELEFAX: (212)564-0525  
 TELEX: 422523  
 INFORMATION FOR SEQ ID NO: 14  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 553 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US94/02252A-14

Query Match 24.4%; Score 55.5; DB 5; Length 553;  
 Best Local Similarity 33.3%; Pred. No. 3.4; Mismatches 7; Indels 5; Gaps 1;  
 Matches 16; Conservative 16; Misnatches 7; Pred. No. 3.4; Indels 5; Gaps 1;

Qy 4 TLSIKNDF---KEIKTDELRLIIGGSGSSTFFRLFNRSFTQALGK 46  
 Db 433 TLRSLGEFDVYQKNISIQDSQVITGNLDISTELGNVNNNISNALNK 480

RESULT 12  
 PCT-US95/10245-13  
 Sequence 13, Application PC/TUS95/10245  
 GENERAL INFORMATION:  
 APPLICANT: SYNTRO CORPORATION  
 TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys And Uses Thereof  
 NUMBER OF SEQUENCES: 60  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: John P. White  
 STREET: 1185 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT-US95/10245  
 FILING DATE: 09-AUG-1995  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: White, John P  
 REGISTRATION NUMBER: 28,678  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212)591-0526  
 TELEX: 422523  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 553 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US95/10245-13

Query Match 24.4%; Score 55.5; DB 5; Length 553;  
 Best Local Similarity 33.3%; Pred. No. 3.4; Mismatches 7; Indels 5; Gaps 1;  
 Matches 16; Conservative 16; Misnatches 7; Pred. No. 3.4; Indels 5; Gaps 1;

Qy 4 TLSIKNDF---KEIKTDELRLIIGGSGSSTFFRLFNRSFTQALGK 46  
 Db 433 TLRSLGEFDVYQKNISIQDSQVITGNLDISTELGNVNNNISNALNK 480

RESULT 13  
 PCT-US95/10245-13  
 Sequence 13, Application PC/TUS95/10245  
 GENERAL INFORMATION:  
 APPLICANT: Bingham, Richard W; Chambers, Philip; Emerson, Peter  
 TITLE OF INVENTION: NEWCASTLE DISEASE VIRUS GENE CLONES  
 NUMBER OF SEQUENCES: 3  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/438,945  
 FILING DATE: 17-NOV-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 885,765  
 FILING DATE: 15-JUL-1986  
 SEQ ID NO:1  
 LENGTH: 553  
 5310678-1

RESULT 14  
 US-09-564-805-228  
 Sequence 228, Application US/09564805  
 GENERAL INFORMATION:  
 APPLICANT: Tavtigian, Sean V.  
 APPLICANT: Teng, David H.F.  
 APPLICANT: Simard, Jacques  
 APPLICANT: Rommens, Johanna M.  
 APPLICANT: Myriad Genetics, Inc.  
 TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility Gene and a Paralog and Orthologous Genes  
 FILE REFERENCE: 2318-258  
 CURRENT APPLICATION NUMBER: US/09/564,805  
 CURRENT FILING DATE: 2000-05-05  
 PRIOR APPLICATION NUMBER: US 60/1107,468  
 PRIOR FILING DATE: 1998-11-06  
 PRIOR APPLICATION NUMBER: 09/434,382  
 PRIOR FILING DATE: 1999-11-05  
 NUMBER OF SEQ ID NOS: 240  
 SOFTWARE: Patentin Ver. 2.0  
 SEQ ID NO: 228  
 LENGTH: 837  
 TYPE: PRT  
 ORGANISM: Arabidopsis thaliana  
 US-09-564-805-228

Query Match 24.4%; Score 55.5; DB 4; Length 837;  
 Best Local Similarity 34.1%; Pred. No. 5.8; Mismatches 9; Indels 9; Gaps 1;  
 Matches 14; Conservative 9; Misnatches 9; Indels 9; Gaps 1;

Qy 7 LKNPKETKTDELRLIIGGSGS-----LSTERFLNR 38  
 Db 473 LPNQLEKIRRDMEEIVLIGGTGSQPSKRNNTSAIFDLFSR 513

RESULT 15  
US-09-242-690A-6  
; Sequence 6, Application US/09242690A  
; Patent No. 6284534  
; GENERAL INFORMATION  
; APPLICANT: KONDO, KEIJI  
; APPLICANT: MIURA, YUTAKA  
; TITLE OF INVENTION: YEAST VECTOR AND METHOD OF PRODUCING PROTEINS USING THE  
; FILE REFERENCE: 049441/0118  
; CURRENT APPLICATION NUMBER: US/09/242, 690A  
; CURRENT FILING DATE: 1999-02-23  
; PRIOR APPLICATION NUMBER: PCT/JP97/02924  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: JP 8/241062  
; PRIOR FILING DATE: 1995-08-23  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Candida utilis  
US-09-242-690A-6

Query Match 23.8%; Score 54; DB 4; Length 97;  
Best Local Similarity 41.4%; Pred. No. 0.59; Matches 6; Mismatches 11; Indels 0; Gaps 0;  
Matches 12; Conservative 12;  
QY 1 MKKTLSLKNDKEKEIKDELEIIGGSSL 29  
Do 43 MKKTYIEENGFRERIKGYEQLYVVASDKL 71

Search completed: July 30, 2002, 15:14:41  
Job time: 517 sec



Query Match 30.4%; Score 69; DB 2; Length 850;  
 Best Local Similarity 33.3%; Pred. No. 1.;  
 Matches 19; Conservative 10; Mismatches 16; Indels 12; Gaps 2;

Qy 2 KKTLSLKNDF----KEIKTBELEIIGGSSLS---TFRRLFNRSTQALGK 46  
 ||| : | | | : ; | : | | | : ||| : | | |  
 ||| : | | | : ; | : | | | : ||| : | | |  
 Db 692 KKTFTTDNVFSFFNLGYTKERKTFNINVSPPGNSQVSESPTRYLDIKTFTEAIQ 748

RESULT 3  
 R52893 hypothetical protein YMR044w - yeast (*Saccharomyces cerevisiae*)  
 N;Alternate names: hypothetical protein YM9532.09  
 C;Species: *Saccharomyces cerevisiae*  
 C;Date: 08-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 05-Nov-1999  
 C;Accession: S52893  
 R;Codel, C.; Bowman, S.  
 submitted to the EMBL data library, February 1995  
 A;Reference number: S52885  
 A;Accession: S52893  
 A;Molecule type: DNA  
 A;Residues: 1-475 <DDE>  
 A;Cross-references: EMBL:248502; NID:9695715; PIDN:CAA88410.1; PID:9695724; GSDB:GN0001  
 C;Genetics:  
 A;Gene: MIPS:YMR044w  
 A;Map position: 13R

RESULT 4  
 H98126 competence stimulating peptide precursor (CSP) [imported] - *Streptococcus pneumoniae* (st  
 C;Species: *Streptococcus pneumoniae*  
 C;Accession: H98126 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
 R;Hoskins, J.A.; Alborn, Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; Delhoff, B.S.; E  
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
 Y.; Sun, P.M.; Winkler, M.E.  
 J. Bacteriol. 183, 5703-5717, 2001  
 A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zock, C.; Baltz, R.H.; Jaskunas, S.R.;  
 A;Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.  
 A;Reference number: A97872; MUID:21429245; PMID:11544234  
 A;Accession: H98126  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-41 <KUR>  
 A;Cross-references: GB:AE007317; PIDN:AAL00845.1; PID:g15459751; GSDB:GN00174  
 C;Genetics:  
 A;Gene: comC

RESULT 5  
 E89792 hypothetical protein SA0272 [imported] - *Staphylococcus aureus* (strain N315)

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

RESULT 6  
 T52611 Query Match 25.6%; Score 58; DB 2; Length 1009;  
 Best Local Similarity 34.3%; Pred. No. 34; Mismatches 8; Indels 0; Gaps 0;  
 Matches 12; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

Qy 9 NDFKEIKTDELEIIGGSSLSITFFRLNRSQTQA 43  
 ||| : | | | : :: : | : | | |  
 ||| : | | | : :: : | : | | |  
 Db 202 NDFPELFDTLVNSISANKDITKWFQTYINKSILSA 236

RESULT 7  
 T52611 Query Match 25.3%; Score 57.5; DB 2; Length 516;  
 Best Local Similarity 41.5%; Pred. No. 19; Mismatches 7; Indels 13; Gaps 5;  
 Matches 22; Conservative 7; Mismatches 11; Indels 13; Gaps 5;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLS--TFF--FRLENRSF 40  
 ||| : | | | : | | | : | | | : | | | : | | : |  
 ||| : | | | : | | | : | | | : | | | : | | : |  
 Db 8 MEKRSTLKNDSFVKYENPIETGSLISIIVGAGSDLAKKTFPALFNLFHQGF 60

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

RESULT 5  
 E89792 hypothetical protein SA0272 [imported] - *Staphylococcus aureus* (strain N315)

C;Species: *Staphylococcus aureus*  
 C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C;Accession: E89792  
 R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O  
 ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K  
 Lancet 357, 1225-1240, 2001  
 A;Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
 A;Reference number: AB9758; MUID:21311952; PMID:11418146  
 A;Accession: E89792  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-1009 <KUR>  
 A;Cross-references: GB:BA000018; PID:913700198; PIDN:BAB41496.1; GSDB:GN00149  
 A;Experimental source: strain N315  
 C;Genetics:  
 A;Gene: SA0272

C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C;Date: 24-Oct-2000 #sequence\_revision 24-Oct-2000 #text\_change 03-Nov-2000  
 C;Accession: T52611  
 R;Wendt, U.K.; Hauschild, R.; Lange, C.; Pietersma, M.; Wenderoth, I.; von Schaewen,  
 Plant Mol. Biol. 40, 487-494, 1999  
 A;Title: Evidence for functional convergence of redox regulation in G6PDH isoforms of  
 plant Mol. Biol. 40, 487-494, 1999  
 A;Reference number: 225284  
 A;Accession: T52611  
 A;Molecule type: translated from GB/EMBL/DDBJ  
 A;Residues: 1-516 <REN>  
 A;Cross-references: EMBL:AJ010970; PIDN:CAB52674.1  
 A;Gene: acq9  
 C;Genetics:  
 C;Keywords: oxidoreductase

Query Match 25.3%; Score 57.5; DB 2; Length 516;  
 Best Local Similarity 41.5%; Pred. No. 19; Mismatches 7; Indels 13; Gaps 5;  
 Matches 22; Conservative 7; Mismatches 11; Indels 13; Gaps 5;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLS--TFF--FRLENRSF 40  
 ||| : | | | : | | | : | | | : | | : |  
 ||| : | | | : | | | : | | | : | | : |  
 Db 8 MEKRSTLKNDSFVKYENPIETGSLISIIVGAGSDLAKKTFPALFNLFHQGF 60

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.83;  
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MKKTLSLKNDFKEIKTDELEIIGGSSLSITFFPLF 36  
 ||| : | : | : | : | : | : | : | : | : |  
 ||| : | : | : | : | : | : | : | : | : |  
 Db 1 MKNTVKLE-QFVALKEKDLOKKGEMRLSKFERP 35

Query Match 25.8%; Score 58.5; DB 2

A;Cross-references: GB:AE009918; PIDN:AAL53790.1; PID:917984720; GSPDB:GN00191	RESULT 10
A;Experimental source: strain 16M	C;Species: <i>Saccharomyces cerevisiae</i>
C;Genetics:	C;Species: <i>Saccharomyces cerevisiae</i>
A;Gene: BMEI0548	C;Alternate names: hypothetical protein L2528
A;Map position: II	C;Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 06-Feb-1998
Query Match 24.9%; Score 55.5; DB 2; Length 398;	C;Accession: S64926
Best Local Similarity 37.2%; Pred. No. 19; Mismatches 7; Indels 1; Gaps 1;	R;Benes, V.; Rechmann, S.; Nentwich, U.; Schwager, C.; Ansorge, W.; Voss, H.
Matches 16; Conservative 5; Mismatches 23; Indels 1; Gaps 1;	A;Reference number: S64920
QY 4 TLISKNDKFKEIKDELEIIGGGS-LSTFFRLFNRSFTQALG 45	A;Accession: S64926
Db 55 TIGVNDNATEDIREGEVFVIMGMSGSGKSTLRLNLRIEPTAG 97	A;Residues: 1-893 <BEN>
RESULT 8	A;Experimental source: strain S288C
D86806	A;Cross-references: SGD:S0004082; MIPS:YLR092W
betaeine ABC transporter ATP binding protein busA [imported] - <i>Lactococcus lactis</i> subsp. <i>lactis</i>	A;Gene: SGD:S0004082
C;Species: <i>Lactococcus lactis</i> subsp. <i>lactis</i>	A;Map position: 12R
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001	C;Keywords: transmembrane protein
C;Accession: D86806	F;136-152/Domain: transmembrane #status predicted <TM1>
R;Boilotin, A.; Wincker, P.; Mauger, S.; Jaiillon, O.; Malarme, K.; Weissenbach, J.; Ehrli, Genom Res. 11, 731-737, 2001	F;223-239/Domain: transmembrane #status predicted <TM2>
A;Title: The complete genome sequence of the lactic acid bacterium <i>Lactococcus lactis</i> subsp. <i>lactis</i>	F;302-318/Domain: transmembrane #status predicted <TM3>
A;Reference number: A86625; MUID:21235186; PMID:1137471	F;358-374/Domain: transmembrane #status predicted <TM5>
A;Accession: D86806	F;411-427/Domain: transmembrane #status predicted <TM6>
A;Status: preliminary	F;488-504/Domain: transmembrane #status predicted <TM7>
A;Molecule type: DNA	F;538-554/Domain: transmembrane #status predicted <TM8>
A;Residues: 1-408 <STO>	F;557-573/Domain: transmembrane #status predicted <TM9>
A;Cross-references: GB:AE005176; PID:912724443; PIDN:AAK05550.1; GSPDB:GN00146	A;Cross-references: SGD:S0004082; MIPS:YLR092W
A;Experimental source: strain IL1403	A;Gene: busA
C;Genetics:	C;Keywords: busA
RESULT 9	Query Match 24.9%; Score 56.5; DB 2; Length 408;
C97848	Best Local Similarity 36.7%; Pred. No. 19; Mismatches 3; Indels 1; Gaps 1;
ABC transporter ATP-binding protein upp [imported] - <i>Rickettsia conorii</i> (strain Malish 7	Matches 16; Conservative 3; Mismatches 15; Indels 1; Gaps 1;
C;Species: <i>Rickettsia conorii</i>	QY 13 EKRUDELEITIGGGS-SLTFFRLFNRSFTQALGK 46
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001	Db 50 EINEGEIFIVIMGMSGKGSTLRLNLRIEPTAG 84
C;Accession: C97848	RESULT 11
R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Roc Science 293, 2093-2098, 2001	Query Match 24.9%; Score 56.5; DB 2; Length 893;
A;Title: Mechanisms of evolution in <i>Rickettsia conorii</i> and <i>Rickettsia prowazekii</i> .	Best Local Similarity 36.7%; Pred. No. 46; Mismatches 18; Indels 7; Gaps 2;
A;Accession: A97700; MUID:21442074; PMID:11557893	Matches 18; Conservative 6; Mismatches 18; Indels 7; Gaps 2;
A;Status: preliminary	QY 1 MKKTSLSKDKFKEITIGGGS-SLTFFRL-----FNRSFTQALGK 43
A;Molecule type: DNA	Db 426 ISKSFRGRINDYKVVPDQL-LAIGVSNLIGFFNAYPATGFSRSALK 473
A;Residues: 1-593 <KUR>	RESULT 11
A;Cross-references: GB:AE006914; PIDN:AAL03725.1; PID:915620316; GSPDB:GN00173	Query Match 24.7%; Score 56; DB 2; Length 719;
C;Genetics:	Best Local Similarity 36.6%; Pred. No. 42; Mismatches 15; Conservative 7; Indels 6; Gaps 1;
A;Gene: upp	Matches 15; Conservative 7; Mismatches 13; Indels 6; Gaps 1;
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology	QY 6 SLKNDFKEIKDELEIIGGGS-SLTFFRLFNRSFTQALGK 40
Query Match 24.9%; Score 56.5; DB 2; Length 593;	Db 37 SLCEQIKELNSDGYEVILVSGAVSAGQRRLRFLKVNSS 77
Best Local Similarity 37.0%; Pred. No. 29; Mismatches 17; Conservative 5; Mismatches 23; Indels 1; Gaps 1;	RESULTS 12
Matches 17; Conservative 5; Mismatches 23; Indels 1; Gaps 1;	E81339
QY 2 KTKLISKNDKFKEIKDELEIIGGGS-LSTFFRLFNRSFTQALGK 46	
Db 298 KNTKLIINNFSPRVRNGEKIGIANGGSKSTFKLKLTKLTPESGK 343	

probable restriction /modification enzyme Cj0690c [imported] - *Campylobacter jejuni* (strain C;Species: *Campylobacter jejuni*)

C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 31-Mar-2000

C;Accession: E81339

R;Parkhill, J.; Wren, B.W.; Mungall, K.; Retley, J.M.; Churcher, C.; Basham, D.; Chillingworth, T.; Quail, M.; Rajandream, M.A.; Rutterford, K.M.; Vaulier, A.; Whitehead, S.; Barrell, B.; Residues: 1-1250 <PAR>

A;Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable regions and cross-references: GB:AL139076; GB:AL111168; NID:96968128; PIDN:CA872964.1; PID:9696814

A;Accession: E81339

A;Status: preliminary

A;Molecule type: DNA

A;Experimental source: serotype O2, strain NCTC 1168

C;Genetics:

A;Gene: Cj0690c

RESULT 13

AC1143 probable NAD(P)-dependent oxidoreductase homolog Imo0546 [imported] - *Listeria monocytogenes*

C;Species: *Listeria monocytogenes*

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C;Accession: AC1143

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, H.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A;Authors: Kraft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitourna, A.; Maok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland, A.; Title: Comparative genomic of *Listeria* species

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AC1143

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-416 <GLA>

A;Cross-references: GB:NC\_003210; PIDN:CA898625.1; PID:g16409922; GSPDB:GN00177

A;Experimental source: strain EGD-e

C;Genetics:

A;Gene: Imo0546

RESULT 14

AF1501 probable NAD(P)-dependent oxidoreductase homolog lin0550 [imported] - *Listeria innocua* (strain C;Species: *Listeria innocua*)

C;Accession: AF1501

C;Accession: AF1501

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, H.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A;Authors: Kraft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitourna, A.; Maok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland, A.; Title: Comparative genomic of *Listeria* species

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AF1501

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-416 <GLA>

A;Cross-references: GB:NC\_003210; PIDN:CA898625.1; PID:g16409922; GSPDB:GN00177

A;Experimental source: strain EGD-e

C;Genetics:

A;Gene: lin0550

Query Match 24.7%; Score 56; DB 2; Length 1250;

Best Local Similarity 40.0%; Pred. No. 77; Matches 14; Conservative 4; Mismatches 15; Indels 2; Gaps 1;

QY 3 KTLSLKNDKEIKDELETTIGGSGSLSFFRLFN 37

Db 380 KNLUSFKRDYKTLSVEHLGTY--EGLISFFETAN 412

**Title:** Comparative genomics of Listeria species  
**Reference number:** AB1077; MUID:2153729; PMID:11679669  
**Accession:** AF1501  
**Status:** Preliminary  
**Molecule type:** DNA  
**Residues:** 1-416 <GLA>  
**Cross-references:** GB:AL592022; PIDN:CAC95782.1; PID:gi16412990; GSPDB:GN00178  
**A: Experimental source:** strain Cip11262  
**C: Genetics:**  
**A: Gene:** lin0550

Wed Jul 31 08:30:53 2002

us-09-833-017-2.rpr



GenCore version 4.5									
Copyright (C) 1993 - 2000 Compugen Ltd.									
OM protein - protein search, using sw model									
Run on: July 30, 2002, 15:14:49 ; Search time 15.86 seconds									(without alignments)
Scoring table: BLOSUM62									12.301 Million cell updates/sec
Database : SwissProt_40:*									
Searched: 105224 seqs, 38779550 residues									
total number of hits satisfying chosen parameters: 105224									
Minimum DB seq length: 0									
Maximum DB seq length: 200000000									
Post-processing: Minimum Match 0%									
Maximum Match 100%									
Listing first 45 summaries									
Result No.   Score   Query   Length   DB ID   Description   Standard   PRT; 475 AA.									
1	60.5	26.7	475	1 YMT4_YEAST	ID YMT4_YEAST	STANDARD;	PRT; 475 AA.		
2	58.5	25.8	41	1 CSP1_SRPN	ID Q04213; YMT4_YEAST	STANDARD;	PRT; 475 AA.		
3	55.9	24.9	893	1 SUL2_YEAST	ID DT 01-NOV-1997 (Rel. 35, last sequence update)	STANDARD;	PRT; 475 AA.		
4	56	24.7	719	1 PSCS_MSCCR	ID DT 01-NOV-1997 (Rel. 35, last annotation update)	STANDARD;	PRT; 475 AA.		
5	55.5	24.4	41	1 CSP_STPR	ID DE HYPOTHETICAL 55.4 kDa protein in MCMI-NUP116 intergenic region.	STANDARD;	PRT; 475 AA.		
6	55.5	24.4	533	1 VGLF_NWVH4	ID OS YMR044W OR YMR0532.09.	STANDARD;	PRT; 475 AA.		
7	55.5	24.4	553	1 VGLF_NDVL	ID OS Saccharomyces cerevisiae (Baker's yeast).	STANDARD;	PRT; 475 AA.		
8	55.5	24.4	553	1 VGLF_NDVTG	ID OC Lukkila, J.; Saccharomyces cerevisiae; Saccharomycetes;	STANDARD;	PRT; 475 AA.		
9	55.5	24.4	553	1 VGLF_NDVTG	ID OC NCBI_TaxID=4932;	STANDARD;	PRT; 475 AA.		
10	55	24.2	608	1 CB22_YEAST	ID RN [1] SEQENCE FROM N.A.	STANDARD;	PRT; 475 AA.		
11	54.5	24.0	373	1 HIS8_MECHJA	ID RP STRAIN=S288C / AB92;	STANDARD;	PRT; 475 AA.		
12	54.5	24.0	463	1 DINA_RICCA	ID RA Odeil, C., Bowman S., Barrell B.G., Rajandream M.A.;	STANDARD;	PRT; 475 AA.		
13	54.5	24.0	720	1 SYFB_PORPU	ID RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.	STANDARD;	PRT; 475 AA.		
14	54	23.8	226	1 CXB2_MOUSE	ID CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ) or send an email to license@isb-sib.ch).	STANDARD;	PRT; 475 AA.		
15	54	23.8	226	1 CXB2 RAT	ID DR EMBL; Z48502; YAA8410.1; -.	STANDARD;	PRT; 475 AA.		
16	54	23.8	708	1 YNBB_CAEEL	ID DR SGD; S000647; YMR044W.	STANDARD;	PRT; 475 AA.		
17	54	23.8	1292	1 RPOC_MCGGE	ID DR InterPro; IPR000313; PWNP.	STANDARD;	PRT; 475 AA.		
18	53.5	23.8	259	1 PSPTB_EDWTB	ID DR Pfam; PF00855; PWNP. 1.	STANDARD;	PRT; 475 AA.		
19	53.5	23.6	553	1 VGLF_NDWD	ID FT DOMAIN 186 195 POLY-GLU.	STANDARD;	PRT; 475 AA.		
20	53.5	23.6	553	1 VGLF_NDUQ	ID FT DOMAIN 243 251 POLY-GLU.	STANDARD;	PRT; 475 AA.		
21	53.5	23.6	553	1 VGLF_NDUV	ID FT SEQUENCE 475 AA; 55426 MW; DDF6189B81178060 CRC64;	STANDARD;	PRT; 475 AA.		
22	53.5	23.6	617	1 CYG2_HUMAN	ID P45972 caenorhabdi	STANDARD;	PRT; 475 AA.		
23	53	23.3	274	1 PPNK_AQKE	ID P4582 mycoplasma	STANDARD;	PRT; 475 AA.		
24	53	23.3	400	1 PROV_ECOLI	ID Q9am14 edwardsiell	STANDARD;	PRT; 475 AA.		
25	53	23.3	400	1 PROV_SALTY	ID P35936 newcastl	STANDARD;	PRT; 475 AA.		
26	53	23.3	513	1 LEU1_LACLA	ID P36515 newcastl	STANDARD;	PRT; 475 AA.		
27	53	23.3	591	1 YM48_YEAST	ID P12270 newcastl	STANDARD;	PRT; 475 AA.		
28	53	23.3	1132	1 DBII_HS62U	ID 073343 homo sapien	STANDARD;	PRT; 475 AA.		
29	53	23.3	1132	1 DBII_HS62Z	ID 067055 aquifex aeo	STANDARD;	PRT; 475 AA.		
30	52.5	23.1	310	1 GTRB_BPP22	ID P57022 bacterioph	STANDARD;	PRT; 475 AA.		
31	52.5	23.1	564	1 Y61A_MCPN	ID P77041 mycoplasma	STANDARD;	PRT; 475 AA.		
32	51.5	22.7	41	1 CSP2_STPR	ID 033690 streptococc	STANDARD;	PRT; 475 AA.		
33	51.5	22.7	240	1 PRKH_MEJJA	ID Q58566 methanococc	STANDARD;	PRT; 475 AA.		

GN COMC1 OR COMC.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; streptococcaceae;  
 OC Streptococcus  
 OC NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-41.  
 RC STRAIN="RX / CP1200";  
 RX MEDLINE=96074663; PubMed=7429953;  
 RA Haeverstein L.-S.; Coomaraswamy G.; Morrison D.A.;  
 RT "An unmodified heptadecapeptide pheromone induces competence for  
 genetic transformation in *Streptococcus pneumoniae*";  
 RT Proc. Natl. Acad. Sci. U.S.A. 92:11140-11144(1995).  
 RN [2]  
 RP STRAIN="RX;  
 RX MEDLINE=97206147; PubMed=9157240;  
 RA Cheng Q.; Campbell B.A.; Naughton A.M.; Johnson S.; Masure H.R.;  
 RT "The com locus controls genetic transformation in *Streptococcus*  
 pneumoniae.";  
 RL Mol. Microbiol. 23:583-592(1997);  
 CC FUNCTION: ACTS AS A PHEROMONE, INDUCES CELLS TO DEVELOP COMPETENCE  
 CC FOR GENETIC TRANSFORMATION.  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- SIMILARITY: BELONGS TO THE COMC FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; U03315; AAC44440 1; -  
 DR InterPro; IPR004218; ComC.  
 DR Pfam; PF03047; ComC\_1.  
 KW Pheromone; Competence.  
 FT PROTEP 1 24 COMPETENCE STIMULATING PEPTIDE TYPE 1.  
 FT CHAIN 1 25 41 AA; 4971 MW; 1448B414E98B86A CRC64;  
 SQ SEQUENCE 41 AA; -----  
 Query Match 25.8%; Score 58.5; DB 1; Length 41;  
 Best Local Similarity 44.4%; Pred. No. 0.34; Indels 1; Gaps 1;  
 Matches 16; Conservative 5; Mismatches 14; Del 1;  
 QY 1 MKKTLSLKDNEKEIKDELEIIGGSGSLSTFRL-----ENRFTQ 36  
 DR ||| : | : | : | : | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 1 MKNTVVKLE-QFVALKEKDLOKKGEMRLSKFFRDF 35  
 -----  
 RESULT 3  
 SUL2\_YEAST STANDARD; PRT; 893 AA.  
 ID SUL2\_YEAST STANDARD; PRT; 893 AA.  
 AC Q12225;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Sulfate permease 2 (High-affinity sulfate transporter 2).  
 GN SUL2 OR SUL2 OR YHR092W OR L9449.1.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetidae; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OC NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.; STRAIN="S288C"; ABT72;  
 RA Johnston M.; Andrews S.; Brinkman R.; Cooper J.; Ding H.; Du Z.;  
 RA Favello A.; Fulton L.; Gatting S.; Greco T.; Kirsten J.; Kucaba T.;  
 RA Hallsworth K.; Hawkins J.; Hillier M.; Johnson D.;  
 RA Johnston L.; Langston Y.; Latrelle P.; Le T.; Mardis E.; Menezes S.;  
 RA

RA Miller N.; Nhan M.; Pauley A.; Peluso D.; Rifkin L.; Riles L.;  
 RA Taich A.; Trzaskos J.; Vignati D.; Wilcox L.; Wohlgemuth P.; Vaudin M.;  
 RA Wilson R.; Waterston R.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.;  
 RA Benes V.; Rechmann S.; Nentwich U.; Schwager C.; Ansorge W.; Voss H.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP CHARACTERIZATION;  
 RA MEDLINE=97201837; PubMed=9055073;  
 RA Cherest H.; Davidian J.C.; Thomas D.; Benes V.; Ansorge W.;  
 RT Surdin-Kerjan Y.;  
 RT "Molecular characterization of two high affinity sulfate transporters  
 in *Saccharomyces cerevisiae*.";  
 RL Genetics 145:627-635(1997);  
 CC -!- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; U53880; AAC67596.1; -  
 DR EMBL; U53865; AAC67550.1; -  
 DR SGD; S0004082; S012.  
 DR InterPro; IPR002645; STAS.  
 DR InterPro; IPR00190; Sulfate\_transp.  
 DR Pfam; PF01140; STAS; 1.  
 DR PROSITE; PS00916; SULFATE\_transp; 1.  
 KW Transport; Transmembrane; Glycoprotein.  
 FT TRANSMEM 132 152 POTENTIAL.  
 FT TRANSMEM 164 184 POTENTIAL.  
 FT TRANSMEM 189 209 POTENTIAL.  
 FT TRANSMEM 222 242 POTENTIAL.  
 FT TRANSMEM 245 265 POTENTIAL.  
 FT TRANSMEM 306 326 POTENTIAL.  
 FT TRANSMEM 351 371 POTENTIAL.  
 FT TRANSMEM 400 420 POTENTIAL.  
 FT TRANSMEM 444 464 POTENTIAL.  
 FT TRANSMEM 484 504 POTENTIAL.  
 FT TRANSMEM 539 559 POTENTIAL.  
 FT TRANSMEM 560 580 POTENTIAL.  
 FT TRANSMEM 893 AA; 99650 MW; 67826955AC7C0BFF5 CRC64;  
 SQ SEQUENCE -----  
 Query Match 24.9%; Score 56.5; DB 1; Length 393;  
 Best Local Similarity 36.7%; Pred. No. 19; Indels 7; Gaps 2;  
 Matches 18; Conservative 6; Mismatches 18; Indels 7; Gaps 2;  
 QY 1 MKKTLSLKDNEKEIKDELEIIGGSGSLSTFRL-----ENRFTQ 43  
 DR 426 ISKSFRGRINDKVVVPDQEL-TAIGVSNLGFFNNPATGGSFSLSALK 473  
 -----  
 RESULT 4  
 PSCS\_MECSR STANDARD; PRT; 719 AA.  
 ID PSCS\_MECSR  
 AC 063361;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Delta-1-pyrroline-5-carboxylate synthetase (PSCS) [Includes: Glutamate  
 5'-kinase (EC 7.2.1.1) (Gamma-glutamyl kinase (GK)); Gamma-glutamyl  
 phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-semialdehyde

DE dehydrogenase) (Glutamyl-gamma-semialdehyde dehydrogenase].  
 GN P5CS.  
 OS Mesembryanthemum crystallinum (Common ice plant).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophytina; Magnoliophyta; eudicots;  
 OC Caryophyllidae; Caryophyllales; Alzoaceae; Mesembryanthemum.  
 NCBI\_TAXID=3544;  
 RN [1]  
 RP SEQUENCE FROM N A.  
 RA Michałowski C.B., Quigley-Landreau F., Bohnert H.J.;  
 RA "Mesembryanthemum crystallinum pyrroline-5-carboxylate synthetase  
 RT mRNA.";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: P5CS PLAYS A KEY ROLE IN PROLINE BIOSYNTHESIS, LEADING  
 CC TO OSMOREGULATION IN PLANTS.  
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate = ADP + L-glutamate 5-  
 CC phosphate.  
 CC -!- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate +  
 CC NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.  
 CC -!- ENZYME REGULATION: FEEDBACK REGULATED BY PROLINE.  
 CC -!- PATHWAY: FIRST AND SECOND STEPS IN PROLINE BIOSYNTHESIS PATHWAY.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN LEAVES AND IS  
 CC INDUCIBLE IN ROOTS SUBJECTED TO SALT STRESS.  
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE GLUTAMATE 5-  
 CC KINASE FAMILY.  
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GAMMA-  
 CC GLUTAMYL PHOSPHATE REDUCTASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the European Bioinformatics Institute and the EMBL outstation -  
 CC use by non-profit Institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; AB067967; AAC18862; 1; -.  
 DR Mendel; 29718; Mscr-1229; 29718.  
 DR InterPro; IPR01046; Akinase.  
 DR InterPro; IPR02486; Aldenzyde\_dehydr.  
 DR InterPro; IPR000565; GPR.  
 DR InterPro; IPR01057; Glut\_5\_kinase.  
 DR Pfam; PF00656; aakinase; 1.  
 DR Pfam; PF00171; aldedh; 1.  
 DR PRINTS; PR00474; GLU5KINASE.  
 DR PROSITE; PS00902; GLUTAMATE\_5\_KINASE; 1.  
 DR PROSITE; PS01223; PRO; 1.  
 KW Proline biosynthesis; Multifunctional enzyme; Oxidoreductase; NADP;  
 KW Transf erase; Kinase.  
 FT DOMAIN 1 293 GLUTAMATE 5-KINASE.  
 FT DOMAIN 294 719 GAMMA-GLUAMYL PHOSPHATE REDUCTASE.  
 FT SEQUENCE 719 AA; 77855 MW; 151011AF559961 CRC64;  
 SQ -----  
 Query Match 24.4%; Score 55.5; DB 1; Length 41;  
 RESULT 6  
 VGLF\_NDVB Best Local Similarity 46.3%; Pred. No. 0.82; Indels 14; Gaps 3;  
 ID VGLF\_NDVB Matches 19; Conservative 5; Mismatches 1; OC  
 AC STANDARD; PRT; 553 AA.  
 DT P06156; 01-JAN-1988 (Rel. 05, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Fusion glycoprotein Precursor [Contains: Fusion glycoprotein F2;  
 DE Fusion glycoprotein Fl].  
 F. GN  
 OS Newcastle disease virus (strain Beaudette C/45) (NDV).  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirus;  
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.  
 OX NCBI\_TAXID=1108;  
 RN [1]  
 RP SEQUENCE FROM N A.  
 RX MEDLINE=8708546; PubMed=3025345;  
 RA Chambers P., Millar N.S., Emerson P.T.;  
 RT "Nucleotide sequence of the gene encoding the fusion glycoprotein of  
 RT Newcastle disease virus.", New Castle disease virus.,  
 RL New Castle disease virus.,  
 RN J. Gen. Virol. 67:2685-2694(1986).  
 RP [2]  
 RP SEQUENCE FROM N A.  
 RX MEDLINE=89204808; PubMed=2705298;  
 RA Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T., Nagai Y.;  
 RA "Newcastle disease virus evolution. II. Lack of gene recombination in  
 DE generating virulent and avirulent strains.", Virology 169:273-282(1990).  
 GN  
 OS Streptococcus oralis.

CC MEMBRANES.  
 CC -!- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2  
 CC LINKED BY A DISULFIDE BOND.  
 CC -!- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X04719; CAA28426.1; -;  
 DR EMBL; M24697; AAA46648.1; -;  
 DR EMBL; A03663; CAA0288.1; -;  
 DR PIR; A27008; VEN2NV.  
 DR HSSP; P04849; ISVF.  
 DR InterPro; IPR00776; Fusion\_gly.  
 DR Pfam; PF00523; fusion\_gly; 1;  
 KW Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;  
 KW Lipoprotein; Palmitate.  
 FT SIGNAL 1 25  
 FT CHAIN 26 553 FUSION GLYCOPROTEIN F0.  
 FT CHAIN 26 116 F2 PROTEIN.  
 FT CHAIN 117 553 F1 PROTEIN.  
 FT TRANSMEM 117 136 POTENTIAL.  
 FT TRANSMEM 136 500 EXTRACELLULAR.  
 FT DOMAIN 137 500 POTENTIAL.  
 FT TRANSMEM 501 527 CYTOPLASMIC.  
 FT DOMAIN 528 553 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 529 585 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 85 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 366 447 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 366 447 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 447 471 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT LIPID 523 523 PALMITATE (POTENTIAL).  
 FT LIPID 523 523 PALMITATE (POTENTIAL).  
 SQ SEQUENCE 553 AA; 59041 MW; 63BF01692AFDE191 CRC64;

Query Match 24.4%; Score 55.5; DB 1; Length 553;  
 Best Local Similarity 33.3%; Pred. No. 15; Matches 16; Conservative 16; Mismatches 7; Indels 20; Gaps 5; Gaps 1; Matches 16; Conservative 16; Mismatches 7; Indels 20; Gaps 5; Gaps 1;

Qy 4 TUSLKNDF---KAIKTELELLIGGSLSLFFRLENRFQITGK 46  
 Db 433 TRLSGEFDVYQKNISIIDSQVITGNLDISTELGNVNNISNALNK 480

RESULT 7  
 VGIF\_NIVH4 STANDARD; PRT; 553 AA.

ID VOLF\_NDV4 STANDARD; PRT; 553 AA.

AC P33613; -;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

Fusion glycoprotein precursor [Contains: Fusion glycoprotein F2; Fusion glycoprotein F1].

DE Fusion glycoprotein F1.

FN Newcastle disease virus (strain Bl-Hitchner/47) (NDV).

OS Paramyxoviridae; Paramyxovirinae; Rubulavirus.

OC Viruses; ssRNA negative-strand viruses; Mononegavirales; Paramyxoviridae; Paramyxovirinae; Rubulavirus.

OX NCBI\_TaxID11181;  
 [1] Sequence from N.A.

RP MEDLINE=89204888; PubMed=2705298;  
 RA Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T., Nagai Y.; "newcastle disease virus evolution. II. Lack of gene recombination in generating virulent and avirulent strains." Virology 169:273-282(1989).

RT -!- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR MEMBRANES.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; M24695; AAA46646.1; -;  
 DR PIR; D46329; DA6329.  
 DR HSSP; P04849; ISVF.  
 DR InterPro; IPR00776; Fusion\_gly.  
 DR Pfam; PF00523; fusion\_gly; 1;  
 KW Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;  
 KW Lipoprotein; Palmitate.  
 FT SIGNAL 1 25 FUSION GLYCOPROTEIN F0.  
 FT CHAIN 26 553 F1 PROTEIN.  
 FT TRANSMEM 117 136 EXTRACELLULAR.  
 FT DOMAIN 501 527 POTENTIAL.  
 FT CARBOHYD 528 553 CYTOPLASMIC.  
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 447 447 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD - 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 553 AA; 58909 MW; FI63E9B7CFDE4493 CRC64;

Query Match 24.4%; Score 55.5; DB 1; Length 553;  
 Best Local Similarity 33.3%; Pred. No. 15; Matches 16; Conservative 16; Mismatches 7; Indels 20; Gaps 5; Gaps 1; Matches 16; Conservative 16; Mismatches 7; Indels 20; Gaps 5; Gaps 1;

Qy 4 TUSLKNDF---KAIKTELELLIGGSLSLFFRLENRFQITGK 46  
 Db 433 TRLSGEFDVYQKNISIIDSQVITGNLDISTELGNVNNISNALNK 480

RESULT 8  
 VGIF\_NIVL STANDARD; PRT; 553 AA.

ID VOLF\_NDV1 STANDARD; PRT; 553 AA.

AC P33614; -;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

Fusion glycoprotein precursor [Contains: Fusion glycoprotein F2; Fusion glycoprotein F1].

FN Newcastle disease virus (strain Las/46) (NDV).

OS Viruses; ssRNA negative-strand viruses; Mononegavirales; Paramyxoviridae; Paramyxovirinae; Rubulavirus.

OX NCBI\_TaxID1114;  
 [1] Sequence from N.A.

RP MEDLINE=89204888; PubMed=2705298;  
 RA Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T., Nagai Y.; "Newcastle disease virus evolution. II. Lack of gene recombination in generating virulent and avirulent strains." Virology 169:273-282(1989).

RT -!- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR MEMBRANES.

CC -----  
 CC -!- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2  
 CC LINKED BY A DISULFIDE BOND.  
 CC -!- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; M24695; AAA46646.1; -;  
 DR PIR; D46329; DA6329.  
 DR HSSP; P04849; ISVF.  
 DR InterPro; IPR00776; Fusion\_gly.  
 DR Pfam; PF00523; fusion\_gly; 1;  
 KW Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;  
 KW Lipoprotein; Palmitate.  
 FT SIGNAL 1 25 FUSION GLYCOPROTEIN F0.  
 FT CHAIN 26 553 F1 PROTEIN.  
 FT TRANSMEM 117 136 EXTRACELLULAR.  
 FT DOMAIN 501 527 POTENTIAL.  
 FT CARBOHYD 528 553 CYTOPLASMIC.  
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 447 447 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD - 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 553 AA; 58909 MW; FI63E9B7CFDE4493 CRC64;

CC -!- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN  
 CC FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; M24696; AAA46647.1; -;  
 DR PIR; B46329; E46329.  
 DR HSSP; P04849; LSVF.  
 DR InterPro; IPR000776; Fusion\_gly.  
 DR Pfam; PF00523; fusion\_gly; 1.  
 DR KW Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;  
 KW Lipoprotein; Palmitate;  
 FT SIGNAL 1 25.  
 FT CHAIN 26 553 | FUSION GLYCOPROTEIN F0.  
 FT CHAIN 26 116 | F2 PROTEIN.  
 FT CHAIN 117 553 | F1 PROTEIN.  
 FT TRANSMEM 117 136 | POTENTIAL.  
 FT DOMAIN 137 500 | EXTRACELLULAR.  
 FT TRANSMEM 501 527 | POTENTIAL.  
 FT DOMAIN 528 553 | CYTOPLASMIC.  
 FT CARBOHYD 85 191 | N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 366 366 | N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 447 447 | N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 471 471 | N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT LIPID 523 523 | PALMITATE (POTENTIAL).  
 SQ SEQUENCE 553 AA; 59047 MW: 04E11F38573B91DF CRC64;

Query Match 24.4%; Score 55.5; DB 1; Length 553;  
 Best Local Similarity 33.3%; Pred. No. 15; DB 1; Length 553;  
 Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

Qy 4 TLSKRNDF---KEIKTDELETTIGGSGSSTFFRLFNRSFTQALGK 46  
 Db 433 TLRISGEFDATYOKNISIQSQVITGNLDISTELGNVNNSISNALNK 480

RESULT 9  
 VGLF\_NDVIG STANDARD; PRT; 553 AA.  
 ID -VGLF\_NDVIG  
 AC P12511;|  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Fusion glycoprotein precursor; [Contains: Fusion glycoprotein F1; Fusion glycoprotein F2];  
 FN Newcastle disease virus (strain Texas 9.b./48) (NDV);  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.  
 RN [1] NCBI\_TaxID=11189;

SEQUENCE FROM N\_A.  
 RX MEDLINE=8826573; PubMed=3388773;  
 RA Schaper U.M., Fuller F.J., Ward M.D.W., Mehrotra Y., Stone H.O.,  
 RA Stripp B.R., de Buyscher E.V.;  
 RT "Nucleotide sequence of the envelope protein genes of a highly  
 RT virulent, neurotropic strain of Newcastle disease virus.";  
 RL Virology 165:291-295(1988).  
 RN [2] SEQUENCE FROM N\_A.  
 RX MEDLINE=8920498; PubMed=2105298;  
 RA Toyota T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T.,  
 RT Nagai Y.;  
 RT "Newcastle disease virus evolution. II. Lack of gene recombination in generating virulent and avirulent strains.";

RL Virology 169:273-282(1989).  
 CC -!- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR  
 CC MEMBRANES.  
 CC  
 CC -!- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2  
 CC LINKED BY A DISULFIDE BOND.  
 CC  
 CC -!- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN  
 CC FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; M23407; AAA46642.1; -;  
 DR PIR; B23201; VGNZGB.  
 DR HSSP; P04849; LSVF.  
 DR InterPro; IPR000776; Fusion\_gly.  
 DR Pfam; PF00523; fusion\_gly; 1.  
 DR KW Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;  
 KW Lipoprotein; Palmitate;  
 FT SIGNAL 1 25.  
 FT CHAIN 26 553 | FUSION GLYCOPROTEIN F0.  
 FT CHAIN 26 116 | F2 PROTEIN.  
 FT CHAIN 117 553 | F1 PROTEIN.  
 FT TRANSMEM 117 136 | POTENTIAL.  
 FT DOMAIN 137 500 | EXTRACELLULAR.  
 FT TRANSMEM 501 527 | POTENTIAL.  
 FT DOMAIN 528 553 | CYTOPLASMIC.  
 FT CARBOHYD 85 191 | N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 366 366 | N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 447 447 | N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 471 471 | N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT LIPID 523 523 | PALMITATE (POTENTIAL).  
 SQ SEQUENCE 553 AA; 59017 MW: 5F537269F876F9E9 CRC64;

Query Match 24.4%; Score 55.5; DB 1; Length 553;  
 Best Local Similarity 33.3%; Pred. No. 15; DB 1; Length 553;  
 Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

Qy 4 TLSKRNDF---KEIKTDELETTIGGSGSSTFFRLFNRSFTQALGK 46  
 Db 433 TLRISGEFDATYOKNISIQSQVITGNLDISTELGNVNNSISNALNK 480

RESULT 10  
 CB32\_YEAST STANDARD; PRT; 608 AA.  
 ID CB32\_YEAST  
 AC B40969;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-OCT-1995 (Rel. 34, Last annotation update)  
 DE Centromere DNA-binding protein complex CBF3 subunit B (Centromere  
 DE protein 3)  
 GN CBF3B OR CEP3 OR YMA18C OR YM8520.17C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccaromycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1] SEQUENCE FROM N\_A.  
 RP STRAIN=AC 502;  
 RX MEDLINE=95045413; PubMed=7957085;  
 RA Lechner J.;  
 RT "A zinc finger protein, essential for chromosome segregation,  
 RT constitutes a putative DNA binding subunit of the Saccharomyces  
 cerevisiae Kinetochore complex, Cbf3.";





Search completed: July 30, 2002, 15:20:52  
Job time: 363 sec

Wed Jul 31 08:30:54 2002

us-09-833-017-2.rsp



Copyright (c) 1993 - 2000 Compugen Ltd.	Gencore version 4.5				
Run on:	July 30, 2002, 15:14:14 ; Search time 45:14 Seconds				
Scoring table:	(without alignments) BLOSUM62 Gapop 10.0 , Gapext 0.5				
Searched:	US-09-833-017-2   562222 seqs, 172994929 residues				
Total number of hits satisfying chosen parameters:	562222				
Minimum DB seqe length:	0				
Maximum DB seqe length:	200000000				
Post-processing:	Minimum Match 0%				
Database :	SPTRMBL19:*				
1:	sp_ archaea:*				
2:	sp_ bacteria:*				
3:	sp_ fungi:*				
4:	sp_ human:*				
5:	sp_ invertebrate:*				
6:	sp_ mammal:*				
7:	sp_ mhc:*				
8:	sp_ organelle:*				
9:	sp_ phage:*				
10:	sp_ plant:*				
11:	sp_ rodent:*				
12:	sp_ virus:*				
13:	sp_ vertebrate:*				
14:	sp_ unclassified:*				
15:	sp_ virus:*				
16:	sp_ bacteriophage:*				
17:	sp_ archaeap:*				
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.					
8	SUMMARIES				
Result No.	Score	Query	Match Length	DB ID	Description
1	227	100.0	46	2	Q99Q15 streptococc
2	220	96.9	46	2	Q9APK7 streptococc
3	209	92.1	43	2	Q9APK6 streptococc
4	69	30.4	850	16	Q97N40 streptococc
5	63.5	28.0	648	2	Q96E51 lactococcus
6	59.5	26.2	275	16	Q9CPPE8 pasteurilla
7	58	25.6	48	2	Q33956 streptococc
8	58	25.6	1009	16	Q9WWU3 staphylococ
9	57.5	25.3	453	3	Q9CIA3 saccharomy
10	57.5	25.3	516	10	Q9SUOK saccharomy
11	57	25.3	516	10	Q91K23 arabidopsis
12	57	25.1	501	2	Q9AII9 oga
13	56.5	24.9	407	2	Q9RQ06 photorhabdu
14	56.5	24.9	407	2	Q9RQ06 lactococcus
15	56.5	24.9	408	16	Q9K1F7 lactococcus
16	56.5	24.9	593	16	Q92GD6 rickettsia
17	56	24.7	1250	16	Q9PP17 campylobact
18	55.5	24.4	125	12	Q41368 newcastle d
19	55.5	24.4	125	12	Q41472 newcastle d
20	55.5	24.4	416	16	Q92EB0 listeria in
21	55.5	24.4	553	12	Q90339 newcastle d
22	55.5	24.4	553	12	Q9DID4 newcastle d
23	55.5	24.4	553	12	Q9RMH7 newcastle d
24	55.5	24.4	553	12	Q92150 newcastle d
25	55.5	24.4	553	12	Q9WE2 newcastle d
26	55.5	24.4	553	12	Q9WIE1 newcastle d
27	55.5	24.4	553	12	Q91Hx4 newcastle d
28	55.5	24.4	553	12	Q91Ab8 newcastle d
29	55.5	24.4	837	10	Q9M819 arabispis
30	55.5	24.4	1304	5	Q9U0H4 plasmid
31	55	24.2	48	2	Q9H56 rickettsia
32	54.5	24.0	230	16	Q9CX0 lactococcus
33	54.5	24.0	255	2	Q9AK13 rickettsia
34	54.5	24.0	255	2	Q9AKN6 rickettsia
35	54.5	24.0	398	16	Q9A1N0 streptococc
36	54.5	24.0	463	16	Q9H56 rickettsia
37	54.5	24.0	819	5	Q97J30 plasmid
38	54	23.8	48	2	Q33651 streptococc
39	54	23.8	235	2	Q9X9T9 vibrio para
40	54	23.8	243	16	Q97m22 clostridium
41	54	23.8	2353	2	P71401 haemophilus
42	53.5	23.6	125	12	Q41386 newcastle d
43	53.5	23.6	125	12	Q41388 newcastle d
44	53.5	23.6	165	1	Q00392 methanococc
45	53.5	23.6	176	2	Q9rj74 streptomyce
RESULT	1				ALIGNMENTS
ID	Q99Q15	PRELIMINARY:	PRT:	46 AA.	
AC	Q99Q15;				
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)				
DT	01-JUN-2001 (TREMBLrel. 19, Last annotation update)				
DE	COMPETENCE STIMULATING PROTEIN PRECURSOR.				
GN	COMC.				
OS	Streptococcus mutans.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; streptococcaceae;				
OX	NCBI_TaxID=1309;				
RN	[1]				
RP	SEQUENCE FROM N. A.				
RC	STRAIN=GB14, HT, LTL1, NC8, AND UAL159;				
RX	MEDLINE=21142515; PubMed=1110878;				
RA	Li Y-H.; Lau P.C.Y.; Lee J.H.; Ellen R.P.; Cvitkovitch D.G.; RT	"Natural Genetic Transformation of Streptococcus mutans Growing in Biofilms,"			
RT	J. Bacteriol. 183:897-908(2001).				
DR	EMBL; AF277152; AAK01543; 1; -				
DR	EMBL; AF277153; AAK01543; 1; -				
DR	EMBL; AF277155; AAK01543; 1; -				
DR	EMBL; AF277156; AAK01546; 1; -				
DR	EMBL; AF277157; AAK01547; 1; -				
DR	INTERPRO; IPR004288; ComC.				
DR	PRAM; PF03047; ComC; 1.				
FT	PRAM; PF03047; ComC; 1.				
SEQUENCE	46 AA:	5211 MW:	38FA62B678FC3BF CRC64;		
8					
Result	Score	Query	Match Length	DB ID	Description
1	227	100.0	46	2	Q99Q15 streptococc
2	220	96.9	46	2	Q9APK7 streptococc
3	209	92.1	43	2	Q9APK6 streptococc
4	69	30.4	850	16	Q97N40 streptococc
5	63.5	28.0	648	2	Q96E51 lactococcus
6	59.5	26.2	275	16	Q9CPPE8 pasteurilla
7	58	25.6	48	2	Q33956 streptococc
8	58	25.6	1009	16	Q9WWU3 staphylococ
9	57.5	25.3	453	3	Q9CIA3 saccharomy
10	57.5	25.3	516	10	Q9SUOK saccharomy
11	57	25.3	516	10	Q91K23 arabidopsis
12	57	25.1	501	2	Q9AII9 oga
13	56.5	24.9	407	2	Q9RQ06 photorhabdu
14	56.5	24.9	407	2	Q9RQ06 lactococcus
15	56.5	24.9	408	16	Q9K1F7 lactococcus
16	56.5	24.9	593	16	Q92GD6 rickettsia
17	56	24.7	1250	16	Q9PP17 campylobact
18	55.5	24.4	125	12	Q41368 newcastle d
19	55.5	24.4	125	12	Q41472 newcastle d
20	55.5	24.4	416	16	Q92EB0 listeria in
21	55.5	24.4	553	12	Q90339 newcastle d
22	55.5	24.4	553	12	Q9DID4 newcastle d
23	55.5	24.4	553	12	Q9RMH7 newcastle d
24	55.5	24.4	553	12	Q92150 newcastle d
25	55.5	24.4	553	12	Q9WE2 newcastle d
26	55.5	24.4	553	12	Q9WIE1 newcastle d
27	55.5	24.4	553	12	Q91Hx4 newcastle d
28	55.5	24.4	837	10	Q9M819 arabispis
29	55.5	24.4	1304	5	Q9U0H4 plasmid
30	55.5	24.2	48	2	Q33615 streptococc
31	55	24.2	303615	16	Q9Cox0 lactococcus
32	54	24.0	819	5	Q97J30 plasmid
33	54	23.8	48	2	Q33651 streptococc
34	54	23.8	235	2	Q9X9T9 vibrio para
35	54	23.8	243	16	Q97m22 clostridium
36	54	23.8	2353	2	P71401 haemophilus
37	54	23.6	125	12	Q41386 newcastle d
38	53.5	23.6	125	12	Q41388 newcastle d
39	53.5	23.6	165	1	Q00392 methanococc
40	53.5	23.6	176	2	Q9rj74 streptomyce

RESULT	2	ID	097N40	PRELIMINARY;	PRT;	850 AA.
O9APK7		AC	097N40;			
		DT	01-OCT-2001 (TREMBLrel. 18, Created)			
		DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
		DR	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
		DE	COMPETENCE STIMULATING PROTEIN.			
		GN	COMC.			
		OS	Streptococcus mutans.			
		OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
		OX	Streptococcus.			
		RN	NCBI_TaxID=1309;			
		RP	[1]	SEQUENCE FROM N.A.		
		RC	STRAIN=BMT1;			
		RX	MEDLINE=21142515; PubMed=11208787;			
		RA	Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.; "Natural Genetic Transformation of <i>Streptococcus mutans</i> Growing in Biofilms."			
		RT	J. Bacteriol. 183:897-908 (2001). ↗			
		RL	EMBL: AF277151; AAC01541.1; - COMPETENCE STIMULATING PROTEIN.			
		FT	CHAIN 26 46 AA; 5195 MW; 38E0B9D5B8RC3BF CRC64;			
		SQ	SEQUENCE 46 AA; 5195 MW; 38E0B9D5B8RC3BF CRC64;			
Query Match	96.9%	Score	220;	DB 2;	Length	46;
Best Local Similarity	97.8%	Pred.	No. 4-2e-21;	Mismatches	1;	Indels
Matches	45;	Conservative	0;	Mismatches	1;	Indels
				Gaps	0;	
RESULT	3	RP	SEQUENCE FROM N.A.			
O9APK6		RC	STRAIN=TIGR4;			
		ID	09APK6	PRELIMINARY;	PRT;	43 AA.
		AC	09APK6;			
		DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
		DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
		DE	COMPETENCE STIMULATING PROTEIN.			
		GN	COMC.			
		OS	Bacteria; Firmicutes; Bacillus/Clostridium group; streptococcaceae;			
		OC	Streptococcus.			
		RC	NCBI_TaxID=1309;			
		RN	[1]	SEQUENCE FROM N.A.		
		RP	STRAIN=JH1005;			
		RX	MEDLINE=21142515; PubMed=11208787;			
		RT	Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.; "Natural Genetic Transformation of <i>Streptococcus mutans</i> Growing in Biofilms."			
		RL	J. Bacteriol. 183:897-908 (2001). ↗			
		DR	EMBL: AF277154; AAC01544.1; - COMPETENCE STIMULATING PROTEIN.			
		FT	CHAIN 26 43 AA; 4927 MW; 65A78FC3BF6156C7 CRC64;			
		SQ	SEQUENCE 43 AA; 4927 MW; 65A78FC3BF6156C7 CRC64;			
Query Match	92.1%	Score	209;	DB 2;	Length	43;
Best Local Similarity	97.7%	Pred.	No. 1e-19;	Mismatches	0;	Indels
Matches	42;	Conservative	1;	Mismatches	0;	Indels
				Gaps	0;	
RESULT	4	RP	SEQUENCE FROM N.A.			
Query Match	92.1%	Score	209;	DB 2;	Length	43;
Best Local Similarity	97.7%	Pred.	No. 1e-19;	Mismatches	0;	Indels
Matches	42;	Conservative	1;	Mismatches	0;	Indels
				Gaps	0;	
RESULT	5	RP	SEQUENCE FROM N.A.			
O9L651		RC	STRAIN=IPLA 972; PLASMID=PBL1;			
		ID	09L651	PRELIMINARY;	PRT;	648 AA.
		AC	09L651;			
		DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
		DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
		DE	HYPOTHETICAL LACTOCOCCIN 972 IMMUNITY PROTEIN (LACTIS PLASMID PBL1 DNA FOR LACTOCOCCIN 972 OPERON).			
		GN	LCLB.			
		OS	Lactococcus lactis (subsp. lactis) ( <i>Streptococcus lactis</i> ).			
		OG	Plasmid PBL1.			
		OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
		RC	Lactococcus.			
		RN	NCBI_TaxID=1360;			
		RP	[1]	SEQUENCE FROM N.A.		
		RC	STRAIN=IPLA 972; PLASMID=PBL1;			
		RX	SANCHEZ C., Hernandez de Rojas A., Martinez B., Arguelles M. E., Suarez J.E., Rodriguez A., Mayo B., "Nucleotide sequence of PBL1, a bacteriocin-producing plasmid from Lactococcus lactis IPLA 972.", Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.			
		RL	[2]	SEQUENCE FROM N.A.		
		RP	STRAIN=IPLA 972; PLASMID=PBL1;			
		RA	Arguelles M. E., Hernandez de Rojas A., Martinez B., Suarez J.E., Rodriguez A., Mayo B., "Nucleotide sequence of PBL1, a bacteriocin-producing plasmid from Lactococcus lactis IPLA 972.", Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.			
		RL	[3]	SEQUENCE FROM N.A.		
		RP	STRAIN=IPLA 972; PLASMID=PBL1;			
		RC	STRAIN=IPLA 972; PLASMID=PBL1;			







RA Weissenbach J.; Ehrlich S.D.; Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 RL lactis ssp. lactis TLL403;"  
 CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
 (ABC TRANSPORTERS).  
 DR EMBL; AF224619; AACF77878.1; -.  
 DR EMBL; AE006375; AAC05550.1; -.  
 DR InterPro; IPR033593; AAA.  
 DR InterPro; IPR003439; ABC\_transportr.  
 DR InterPro; IPR001687; ATP\_GTP\_A.  
 DR InterPro; IPR000644; CBS.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR SMART; SM00382; AAA; 1.  
 SMART; SM00116; CBS; 2.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 KW ATP-binding; Complete proteome; Transport;  
 SEQUENCE 408 AA; 45750 MW; EBBA8A87BBS563B88 CRC64;

Query Match 24.9%; Score 56.5; DB 16; Length 408;  
 Best Local Similarity 45.7%; Pred No 46;  
 Matches 16; Conservative 3; Mismatches 15; Indels 1; Gaps 1;  
 QY 13 EIKTDLLEITIGGSGS-LSPFLRFLNRSFTQALGK 46  
 Db 50 EINEGEIIFVINGLSGSGKSTLRLNRLNRIEPPSGK 84

Search completed: July 30, 2002, 15:20:30  
 Job time: 376 sec

GenCore version 4.5  
 Copyright (c) 1993 - 2000 Compugen Ltd.

**OM protein - protein search, using sw model**

Run on: July 30, 2002, 15:14:11 ; Search time 53.42 Seconds  
 (without alignments)  
 916.951 Million cell updates/sec

Title: US-09-833-017-4  
 Perfect score: 2173  
 Sequence: MNEALMILSNGLLITYLVLF.....KTSLOTSNHHHLFKQLLKK 441.

**Scoring table:** BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
 Maximum DB seq length: 200000000 |

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

**Database :**

A\_Geneseq\_032802;\*  
 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT;\*  
 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT;\*  
 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT;\*  
 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT;\*  
 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT;\*  
 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT;\*  
 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT;\*  
 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT;\*  
 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT;\*  
 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT;\*  
 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT;\*  
 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT;\*  
 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT;\*  
 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT;\*  
 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT;\*  
 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT;\*  
 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT;\*  
 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT;\*  
 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT;\*  
 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT;\*  
 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT;\*

Pre. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query Match Length	DB ID	Description
1	752	34.6	446	20 AAY14065
2	752	34.6	446	20 AAY103650
3	752	34.6	446	20 AAW89437
4	752	34.6	446	20 AAW89434
5	398	18.3	175	19 AAY85911
6	378.5	17.4	432	22 AAG81813
7	258	11.9	108	19 AAW38531
8	192	8.8	106	19 AAW38330
9	189	8.7	106	18 AAY1170
10	151	6.9	543	22 AAG98391
11	145.5	6.7	442	20 AAW83369

**RESULT** 1  
 ID AAY14065 standard; Protein: 446 AA.  
 XX AAY14065;  
 XX 16-JUL-1999 (first entry)

DE S . pneumoniae Ornithine carbamoyltransferase.  
 XX KW Ornithine carbamoyltransferase; argF/arcB family polypeptide; diagnosis;  
 KW infection; therapy; immune response; otitis media; conjunctivitis;  
 KW pneumonia; bacteraemia; sinusitis; pleural empyema; endocarditis;  
 KW meningitis; matrix protein adhesion.

OS Streptococcus pneumoniae.  
 XX EP913475-A2.  
 PN XX EP913475-A2.  
 PD XX 06-MAY-1999.  
 PF XX 22-OCT-1998; 98EP-0203571.  
 PR XX 30-OCT-1997; 97US-0961536.  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX PI Brown JR, Zalacain M;  
 XX DR WPI: 1999-256630/22.  
 DR N-PSDB; AAY57966.  
 XX PT New Streptococcus pneumoniae ornithine carbamoyltransferase; argF/arcB polypeptide and polynucleotide, usp

PT antibacterial drugs, and for diagnosis and treatment of streptococci  
infections which cause meningitis etc

XX

PS

Disclosure: Page 27-28; 30pp; English.

XX

CC This sequence is the streptococcus pneumoniae ornithine carbamoyltransferase protein (I) of the invention. The ornithine carbamoyltransferase protein is a (arg/arcB) family polypeptide. (I) and poly nucleotides (II) encoding it are useful for diagnosing diseases due to an infection of an organism with the ornithine carbamoyltransferase gene by determining the nucleic acid sequence encoding (I), and/or analysing for the presence or amount of (I). They can diagnose the stage and type of infection. (I) is also useful for screening for compounds which affect activity of the protein by measuring the binding to (I) and observing the stimulation or inhibition of the polypeptide function. These can be used in treatment to inhibit or enhance ornithine carbamoyltransferase activity, in addition to direct administration of ornithine carbamoyltransferase polypeptides to treat conditions associated with a lack of ornithine carbamoyltransferase polypeptide, or direct administration of antisense sequences to prevent expression. (I) and antibodies against it induce an immune response to immunise and prevent disease. Anti-(I) antibodies induced by the polypeptide are also useful for isolating clones expressing (I), or for purifying the polypeptide by affinity chromatography. Diseases diagnosed, prevented or treated include otitis media, conjunctivitis, pneumonia, bacteræmia, sinusitis, pleural empyema, endocarditis and especially meningitis. (I), (II) and their antagonists can prevent adhesion of bacteria to matrix proteins and are useful for use on wounds and body implants to prevent bacterial infection.

CC

XX

Sequence

446 AA;

SQ

PT

DE

XX

AC

XX

DT

XX

DE

XX

S. pneumoniae histidine kinase polypeptide.

XX

KW

Histidine kinase; two component signal transduction system; TCSTS;

CC

bacterial; immune response; infection; otitis media; conjunctivitis;

CC

bacteræmia; sinusitis; Pleural empyema; endocarditis; meningitis;

CC

bacterial adhesion; wound.

OS

Streptococcus pneumoniae.

XX

PN

EP892059-A2.

XX

PD

20-JAN-1999.

XX

PF

17-JUN-1998;

98EP-0304782.

XX

PR

20-JUN-1997;

97US-0879941.

XX

(SMIK ) SMITHKLINE BEECHAM CORP.

(SMIK ) SMITHKLINE BEECHAM PLC.

XX

PI

Wallis NG;

XX

DR

WPI; 1999-083556/08.

XX

N-PSDB; AAX29247.

XX

PT

New Histidine Kinase polypeptide and poly nucleotide - useful as

PT

diagnostic reagents and for prevention and treatment of

PT

Streptococcus pneumoniae infections, especially meningitis

XX

PS

Claim 6; Page 9; 37pp; English.

XX

CC

This

represents

a histidine kinase (HK) polypeptide which is a component

CC

of the two component signal transduction system (TCSTS) in bacteria. The

CC

gene encoding the mature polypeptide can be isolated from S. pneumoniae

CC

0100931. Host cells containing a vector comprising the nucleic acid are

CC

used for the recombinant expression of the protein. HK Polypeptides and

CC

polynucleotides are useful for diagnosing diseases related to over

CC

or underexpression of HK protein by identifying mutations in the HK gene, or

CC

determining HK polypeptide or mRNA expression levels due to an infection

CC

of an organism with the HK gene. The polypeptides are also useful for

CC

screening for compounds which affect activity of the protein. These can

CC

be used in treatment to inhibit (antagonist i.e. antibiotic drugs) or

CC

enhance (agonist) HK activity. The polypeptides (administered directly,

CC

in a vector and as a vaccine) and antibodies induce an immune response to

CC

immunise and prevent disease. Diseases diagnosed, prevented or treated

CC

include: bacterial infections, especially S. pneumoniae infections, which

CC

cause otitis media, conjunctivitis, bacteræmia, sinusitis, pleural

CC

empyema, endocarditis and especially meningitis. The polypeptides,

CC

polynucleotides and their (antagonists can prevent adhesion of bacteria to

CC

polymyxins; and are useful for use on wounds and body implants to

CC

prevent bacterial infection.

XX

Sequence

446 AA;

SQ

Query Match

34 6%

Score

752;

DB

20:

Length

446:

Best Local Similarity

39 2%

Pred No

3e-58;

Mismatches

169;

Indels

12;

Gaps

5;

Matches

172;

Conservative

86;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conservat

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;

Mis

atches

169;

Indel

s

5;

Gap

5;

Match

172;

Conserv

e;









PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in  
 PT vaccines, drug screening, etc  
 XX  
 PS claim 6; Page 226; 354pp; English.  
 XX  
 CC AAX30724 to AAX30946 represent genomic DNA sequences isolated from  
 CC Streptococcus pneumoniae strain 010093. These genomic DNA sequences  
 CC encode the novel proteins given in AAY1114 to AAY11367. The proteins,  
 CC isolated from Streptococcus pneumoniae, can be used in vaccines against  
 CC streptococcal infections and in assays for identifying compounds that  
 CC inhibit or activate the activity of the proteins. The antagonists can  
 CC be used to treat an individual having need to inhibit a bacterial  
 CC protein. Vectors expressing the proteins can be used to induce a  
 CC protective immune response in mammals.

XX Sequence 106 AA;

SQ

Query Match 8.7%; Score 189; DB 18; Length 106;  
 Best Local Similarity 40.7%; Pred. No. 3.5e-09; Mismatches 37; Conservative 22; Indels 0; Gaps 0;  
 Matches 37; Mismatches 32; Indels 0; Gaps 0;  
 OY 170 YILLIOLVLYWFSYNSVNPILKFRKEVWVYLILILISFLSQTOKVQNEMMAKREAQ 229  
 Db 4 YILVMDLSIYVEEYIGLQSTVTRVHLIVLIFMmgikkidykeklqeeningdtlr 63  
 OY 230 IINITYOSQSQESTLKDIDSPRHDVNTS 260  
 Db 64 Yrdmerysrhieelykeirsfrhdynltt 94

RESULT 10

AC AAG98391

ID AAG98391 standard; Protein: 543 AA.

XX

AC AAG98391;

XX

DT 21-SEP-2001 (first entry)

XX

Escherichia coli protein sequence SEQ ID NO:439.

XX

KW Escherichia coli; identification; proliferation; microorganism;

KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;

KW bacterial growth inhibition.

XX

OS Escherichia coli.

XX

PN WO200148209-A2.

XX

PD 05-JUL-2001.

XX

PF 19-DEC-2000; 2000WO-US34419.

XX

PR 23-DEC-1999; 99US-0173005.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Forsyth RA, Ohlsen KL, Zyskind JW;

XX

DR WPI; 2001-4537649.

XX

N PSDB; AABH1447.

XX

X Novel nucleic acids encoding proteins required for Escherichia coli

X proliferation, useful for screening for antimicrobial agents.

X Claim 19; Page 558-559; 596pp; English.

XX

CC The present invention describes a purified or isolated nucleic acid

CC sequence (I) consisting essentially of one of the 93 nucleotide sequences

CC given in AAB11202 to AAB11294, where expression of the nucleic acid in a

CC microorganism is capable of inhibiting proliferation of a microorganism.

CC (I) have antibiotic and antibiotic activities, and can be used in

CC gene therapy. Expression of (I) in a microorganism inhibits proliferation

XX

CC of the microorganism, and the manufactured antibiotic is useful for  
 CC reducing the activity or level of a gene product required for  
 CC proliferation of a microorganism in a subject, specifically humans. The  
 CC nucleic acids that inhibit bacterial growth or proliferation can be used  
 CC as antisense therapeutics for killing bacteria. In addition to  
 CC therapeutic applications, the nucleic acid sequences complementary to  
 CC sequences required for proliferation can be used as diagnostic tools.  
 CC For example, nucleic acid probes complementary to proliferation-required  
 CC sequences that are specific for particular species of microorganisms can  
 CC be used as probes to identify particular microorganism species in  
 CC clinical specimens. AAB11205 to AAB11487 encode the Escherichia coli  
 CC proteins given in AAG98239 to AAG9831, and AAB11488 to AAB11491  
 CC represent oligonucleotides, which are used in the exemplification of the  
 CC present invention.

XX Sequence 543 AA;

SQ

Query Match 6.9%; Score 151; DB 22; Length 543;  
 Best Local Similarity 23.4%; Pred. No. 8.2e-05; Mismatches 89; Conservative 68; Indels 80; Gaps 18;  
 Matches 89; Mismatches 143; Indels 80; Gaps 18;

OY 85 IFYGLIPVASSDFF--RRAIFFFILDGTQDGIVMGSSITTMIEFAGIALSYL-----P 136

Db 208 lfqlepeystifeqrqamlsiqvavdargev-----tindaqellnyrkssddk 265

OY 137 LSVN-----VDIGR-LKDSDLT---MKVKKRLIPPMNTMLVYLLIQVLYIESINV 185

Db 266 lstlshswsqvdsevirgdgprdrdebeitikdrllin-----304

OY 186 IPTLKFRKVWVYLILFLILISFLSQTOKVQNEMMAKREAQ 245

Db 305 --tvpyrsngi-----1gaistf-----1gatvkrkingridgvnyad 342

OY 246 DIRSFHDYLNIELTSLRIGENKLASIEKLYHQILEKTGHQLQDTRINIGH-LANIQD 304

Db 343 alrershefmnkhlvi-1qllh--1ksyqkedyil-ktanyqe---eisglgiksp 395

OY 305 AVKGILSAKILEAQNKTKAVNEVSSKIQPEMELDFILISLICDNATEAESLNP- 363

Db 396 viagfliskinratd--lightlinssqdsgsdqvatlittignienalealge 453

OY 364 --EQLARFKKNSAVIPIQNSTKEQID-VSKIREVNSTKGNSRGIAKRNILE 419

Db 454 pggeslvthylhyrigwlhevdngpgsapdkidhifokgystkgsergyqlalvkkqven 513

OY 420 YPKTLQTSNNHHLFKQMLI 439

Db 514 1-ggsiavsesepgqifqfv 532

RESULT 11

AAW83369

ID AAW83369 standard; Protein: 442 AA.

XX

AC AAW83369;

XX

DT 17-FEB-1999 (first entry)

XX

DE Streptococcus pneumoniae histidine kinase.

XX

KW Streptococcus pneumoniae; histidine kinase; antibacterial; diagnosis;

KW streptococcal; Helicobacter pylori; infection; bacteriostatic; microbial;

KW bacteriocidal; ulcer; gastritis; stomach cancer; gene therapy.

XX

OS streptococcus pneumoniae.

XX

PN EP881286-A2.

XX

PD 02-DBC-1998.

XX

PF 26-MAY-1998; 98EP-0304138.

XX

PR 30-MAY-1997; 97US-0048347.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Biswas S, Throup J, Wallis NG, Zalacain M;  
 XX DR WPI; 1999-001392/01.  
 XX DR N-PSDB; AAV72649.

XX PT New Streptococcus pneumoniae Histidine kinase polypeptide and  
 PT and treatment of Streptococcal and Helicobacter pylori infections  
 XX PS Claim 1; Page 34-35; 44pp; English.

XX The present sequence represents histidine kinase (HK) isolated from  
 CC Streptococcus pneumoniae. HK polynucleotides and polypeptides are useful  
 CC for diagnosing susceptibility to diseases by detecting mutations or  
 CC polymorphisms in the HK gene or analysing for the presence or amount  
 CC of HK polypeptide expressed in a patient sample. HK PCR probes are  
 CC useful for diagnosing diseases (especially Streptococcal), and can  
 CC characterise the stage and the species or strain causing the infection.  
 CC The HK probes can also determine the response of the infectious organism  
 CC to drugs. HK polypeptides and polynucleotides are useful for screening  
 CC for antagonists, agonists and drugs against infectious micro-organisms.  
 CC HK agonists and antagonists are bacteriostatic and bacteriocidal  
 CC compounds which can be used in treatment to enhance (agonist) or block  
 CC (antagonist or antisense sequence) HK activity, therefore treating  
 CC microbial (especially Streptococcal) diseases, ulcers and gastritis, and  
 CC stomach cancer caused by Helicobacter pylori. Epitopes of HK polypeptides  
 CC and polynucleotides are useful immunogens for producing anti-HK  
 CC antibodies for vaccines to prevent bacterial infections, and HK  
 CC polynucleotides can be used in genetic immunisation (gene therapy) to  
 CC prevent infections. HK polypeptides and polynucleotides and their  
 CC (antagonists can prevent adhesion of bacteria to matrix proteins, and  
 CC are useful for use on wounds and body implants to prevent bacterial  
 CC infection.

XX Sequence 442 AA;

XX SQ Query Match 6.7%; Score 145.5; DB 20; Length 442;  
 CC Best Local Similarity 21.7%; Pred. No. 0.00019; Mismatches 154; Indels 151; Gaps 21;  
 CC Matches 104; Conservative 70; Mismatches 154; Indels 151; Gaps 21;

QY 24 LFSKISNVTLSSKKELTFLSISINFLIMIAVTMWNVLFLYPAAEFLYALSTYL-NRONSLS 82  
 Db 6 lftkkiyfrs-----ifsvlviichhai-----yfilpstylyrqtig 46

QY 83 LNIFIGLPLPVASSOLFRRRIIFILFD-----GTOGIVMGSSIIITYMEFAGIALSYLF 136  
 Db 47 qkataiqslaqslqsgd-----rqsiqvldysqtsdikgtvke-----mte----- 87

QY 137 LSVENVDIGRKDSL-----TKMKVKKRLLPMNT-TMLYYILL-----QVIX 178  
 Db 88 -----dklevkdsp1pltdqtslflfeerektqdggtmlqlfasmalqkaeqsl 141

QY 179 VIESYNVPIPKFRKVWVYLLFLUINSFSQYTKVKVNEIMAKEAQIR----- 231  
 Db 142 qfplytiasflsilvvayiyartivapilekrtr-----mndlsqvrirvdskd 195

QY 232 NITRQSQISITSYK-----TFRSFIDYLNTTSURGE 266  
 Db 196 eignlkgqinlyqghltviadheknealglekmkvfingashelktpaslkile 255

QY 267 N-----KD-----LASIEKIVHOLE-----KTGHQDQDPRYNHTGLANTQNDAV 306  
 Db 256 nnrenigrykrdqylgvalgivedlnhvhqilisvsqellddrddtildqntqnlv 315

QY 307 KQILSKILKEQ-----NKKIAVNEVSSKQLPPEMELLPTTISILCDNAEAFES 350  
 Db 316 dyallakkerelqiduslthqgqaylpsv-----mk1-----ilsnisnaikhsv-- 360

XX PT New nucleic acid and peptides, useful as antibiotic peptides -  
 XX PS Claim 38; Page 128-130; 151pp; English.

CC The present invention describes peptide antibiotics which inhibit the  
 CC growth of pneumococci, ABC transporter and two-component signal  
 CC transduction system proteins from Streptococcus pneumoniae. The peptides  
 CC are useful (especially in the form of a pharmaceutical composition) for  
 CC the treatment of a bacterial infection or inflammation. Methods from the  
 CC present invention are also useful for the identification of agents or  
 CC drugs which are useful in preventing bacterial proliferation or kill  
 CC bacterial cells e.g. potent antibiotics. The peptides and methods are  
 CC also useful for providing important epidemiological tools. The peptide  
 CC are especially useful for the prevention of any disease caused by a  
 CC bacterium e.g. Staphylococcus aureus, Acinetobacter, Enterococcus  
 CC faecalis, Escherichia coli, Pseudomonas aeruginosa all of which can  
 CC cause blood poisoning among other ailments, Mycobacterium tuberculosis  
 CC which causes tuberculosis, Shigella dysenteriae which causes dysentery,  
 CC and Neisseria gonorrhoeae which causes gonorrhoea. The peptides are also  
 CC useful in the treatment of infections due to Streptococcus pneumoniae a  
 CC bacterial species that causes blood poisoning, middle ear infections,  
 CC pneumonia and meningitis in humans. The present sequence represents a  
 CC sensor histidine kinase from the present invention.

XX Sequence 442 AA;

XX SQ Query Match 6.7%; Score 145.5; DB 21; Length 442;  
 CC Best Local Similarity 21.7%; Pred. No. 0.00019; Mismatches 154; Indels 151; Gaps 21;  
 CC Matches 104; Conservative 70; Mismatches 154; Indels 151; Gaps 21;

QY 24 LFSKISNVTLSSKKELTFLSISINFLIMIAVTMWNVLFLYPAAEFLYALSTYL-NRONSLS 82  
 Db 6 lftkkiyfrs-----ifsvlviichhai-----yfilpstylyrqtig 46

QY 83 LNIFIGLPLPVASSOLFRRRIIFILFD-----GTOGIVMGSSIIITYMEFAGIALSYLF 136  
 Db 47 qkataiqslaqslqsgd-----rqsiqvldysqtsdikgtvke-----mte----- 87

QY 137 LSVFNVDIGRLKDSL-----TKMKVKRKLIPWMNITMMLVYLLI-----OVLV 178

DB 88 -----dklevkslpltdqrlqtsifereevktqdggmmlqkaedsl 141

QY 179 VESYNVPIPTKFRKEVWIVLILFLISFLSQTOKQVNEMAKREAQR----- 231

DB 142 qfipytilasflislvayiyartavipalektrvtrr-----mndidsqvrlrvskd 195

QY 232 NTQYSQQESIYKD-----IRSFHDYLNLTSLRGIE 266

DB 196 eignlkeginslyghlvtiadlhknealqlmekmkvfirgasheiktpiaskilie 255

QY 267 N-----KD-----LASIEKTYHQLE--KGHOLDTRYKNIGHLANIQDAV 306

DB 256 nnrenigrykardqylgvalgvialhvlqlsllsvqelldqntqnvk 315

QY 307 KSLSKAKILEAQ-----NKIAVNEVSSKIQPEMELDFITLSICDNAAEES 360

DB 316 dyallkerelqidnsltbhqayiipsv-----mk1-----ilsnlshnaihsv-- 360

QY 361 LNEPIOLIAFFKKNGSIVFITIONSTEKEOIDVSKIFKENYSTKSNSNRGIGLAKVNTHLE 419

DB 361 --pggivrigregej-fienscseeqeklaqsfdsnasrkvgsgmgfvvksleh 416

RESULT 13

AAW69977 standard; Protein; 861 AA.

XX AAW69977;

AC AAW69977;

DT 19-NOV-1998 (first entry)

DE Staphylococcus aureus Histidine kinase.

XX Histidine kinase; KadP; prevention; treatment; diagnosis; vaccine.

XX Staphylococcus aureus.

OS EP863208-A2.

XX 09-SEP-1998.

PD 17-FEB-1998; 98EP-0301167.

PR XX 25-FEB-1997; 97US-0039478.

PA (SMK ) SMITHKLINE BRECHAM CORP.

PA (SMK ) SMITHKLINE BEECHAM PLC.

XX Wallis NG;

DR WPT; 1998-4588394/40.

XX N-PSDB; AAV43778.

PT 1998-4588394/40.

XX WPT; 1998-4588394/40.

DR AAV43778.

XX WPT; 1998-4588394/40.

DR AAV43778.

PT 1998-4588394/40.

XX New DNA encoding Staphylococcus aureus histidine kinase used to prevent, treat, diagnose and vaccinate against - e.g. respiratory tract infections and cardiac, gastrointestinal, CNS, eye, kidney, urinary tract, skin, bone and joint disorders

XX X claim 14; Fig 2; 31pp; English.

XX The present sequence represents a Staphylococcus aureus Histidine kinase.

CC The protein is related, by amino acid homology, to the Escherichia coli Kdp histidine kinase protein. The protein can be used to treat an individual in need of histidine kinase, while the antagonist is used to inhibit histidine kinase. DNA encoding and expressing histidine kinase are used in immunological compositions. The active agents are useful for preventing, treating, diagnosing and vaccinating against infections of the upper and lower respiratory tract, cardiac disorders, gastrointestinal disorders, eye disorders, kidney and urinary tract disorders, skin disorders, and bone and joint disorders.

CC They can also be used for assaying genetic variation and raising an immunological response against e.g. Staphylococcus aureus.

XX

SQ Sequence 861 AA;

Query Match 6.6%; Score 143; DB 19; Length 861; Best Local Similarity 21.2%; Pred. No. 0.00081; Matches 109; Conservative 79; Mismatches 155; Indels 172; Gaps 25;

DB 414 tilimifilq-----illsiwtisfigflaaininvfnhyffeptryfev 462

QY 65 ----PLVLFLLFLFSRSVNTLSKELTFLSI--SNFLIMIAVTMWNLF-----IPAE-- 64

DB 463 rfdypitfi-vsitsiltsalkqikfqysikkqlyrtalifqind-----sik 512

QY 121 TTYMEFAG-TALSYLFLSVFNVNDIGRLKDSLTKMVKKRKLIPWMNITMMLVYLLQVLYI 180

DB 513 qtytvenlinanagydi-----nqlqgsit-----iyvi 541

QY 181 ESYNVIPTKFRKEV-----VIVL-----LFLLIS----- 208

DB 542 ngskvikptlqnhdinttqhgqealswvknernqagattdfpqinkwlipgspikg 601

QY 209 -FLSOYTQVON-----EIMAQERQKRITQIQSQQTESLYKD-----IRSF 250

DB 602 laidayqssqvnpdydasilesmineislaivenvtlikqgresmlqaergqlthsnflrsi 661

QY 251 RHDYLNILISIRLGIE-----NKDLASEKTYHQLEKTGHQLODTRY-----NIGHLA 299

DB 662 shdirtplttmgldilvshksdmisile-kqly-----hsfqsqlylvtnhsil 716

QY 300 NIQNDAVKGTLSA-----KILERQKKIAVNEVSSKIQPEMELDFITLSI 348

DB 717 klqssvqklqpylvselweedmiledrhkrkrivssvnlqfhidsklqaln 776

QY 349 LCDNATE-AFAESLNPEIQLAFFFKNNSIVFITION-----STKEOIDVSKIFKENYST 401

DB 777 lienavkhststdtklnslysiyasq---iefavidepgqisleeqq---kifepry- 828

QY 402 KGSNR-----GIGAKVNTHLEMPKTLQ 426

DB 829 -gsnkypkdkqkesmglglylvqtihkhqsnlq 861

RESULT 14

AAU37397 standard; Protein; 909 AA.

XX ID AAU37397

XX AC AAU37397;

XX DT 14-FEB-2002 (first entry)

XX DE Staphylococcus aureus cellular proliferation protein #1567.

XX KW Antisense; prokaryotic cellular proliferation protein;

XX KW antibiotic; antibacterial; drug design.

XX OS Staphylococcus aureus.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PR 21-MAR-2001; 2001WO-US09180.

XX PR 23-MAY-2000; 2000US-206848P.

XX PR 26-MAY-2000; 2000US-207727P.

XX PR 23-OCT-2000; 2000US-243578P.

XX PR 27-NOV-2000; 2000US-253625P.



	Best Local Similarity	19.7%	Pred.	No. 0.00027;
Matches	84;	Conservative	83;	Mismatches
QY	KEULLFSNLFMLI-ATVMMVNLFYPAEPLFIASIYLRLRQNSSLNITYGLIPVAS	94	Indels	135; Gaps
Db	5 kdtssaeisslilylpiagifhevaygpkwlylisvivt----sis-----yilivin	55		
Qy	SDLERRATIFILDGTCIGMGSISITYMIEPAGIALSYLSLVSFNVDIGRKSITKM	154		
Db	56 nri--ntimfyil----lihyficiyfvfsvhpmislfyfsafavpf-tfknnvkt	107		
Qy	KVKRKLIPMNTMULYLLQLVLYIESYNVIPLKFRKKVVVYLFLILFISLSQT	214		
Db	108 atnifnlitkicdityllynyfva-----mnyyyvvvslinlnfkmk	153		
Qy	KOKYQNETMAQKEAQIRNITYOQSIESLYKDITRSFRHDYLNLTSRLG-TENKDLASIE	274		
Db	154 nreypkel-aekorhintl-----	171		
Qy	KIVHOILEKTGHQLOQDTRYNIGHIANIONDAVGILSAKLERONKKI---AVNV-	326		
Db	172 -iaequerhingqdhat--igny-faslslkselayskliaddevkvkaellainklsr	225		
Qy	327 -----EVSKOLPE-MELLFRITISILCNOAIAFAESUNPELOLAFFKKNGSIVE	378		
Db	226 eslkvkreiddavkplspid---sirkvlkdadifte--nkealaqvspktkgmlv	281		
Qy	379 IIQNSTKEQIDVSKIFKENYSTKGNSNRGIGLAKVNHLIEHPKT---SLOTSNHHLF	434		
Db	282 mi---tre-----ainnvikhanaskvhgklikvnhh---310			
Qy	435 KQLIIK 441			
Db	311 killimie 317			

Search completed: July 30, 2002, 15:14:13  
Job time: 1594 sec



Sequence 2, APPI1  
Sequence 2, APPI1  
Sequence 22, APPI1  
Sequence 1, APPI1  
Sequence 16, APPI1  
Sequence 1, APPI1  
Sequence 4, APPI1  
Sequence 4, APPI1  
Sequence 2, APPI1  
Sequence 8, APPI1  
Sequence 6, APPI1  
Sequence 6, APPI1  
Sequence 6, APPI1  
Sequence 6, APPI1  
Sequence 4, APPI1  
Sequence 4, APPI1  
Sequence 4, APPI1

Db 6 ILYALVINGLKIVIFFKUNGIGLEFDRIFKAFLKFLGLITTEQFLAVSKYLSFIE 65  
 Qy 65 PYFLTAISIYURQNSLNLNFFYGLPVAASSDERRATIFTDGT-QGIV-MGSII 121  
 Db 66 PFLFGILSFLLRLGPKKLILYGFPMILVEYRGYSYEVPLQGIVDGNPFL 125  
 Qy 122 TYMEFAGIALSYLFSVENNDIGRKLDSLTKMKVKKRLLPMMTMLYLLQVWIE 181  
 Db 126 LIMIFVCIFL-VFLKWDYDTRREFLDTGFOQLSUTKINWAMGAYLWMSYLE 183  
 Qy 182 SYNVIPTLKFRKVVTYLILFLILTSLSQYTKQVKVNEIMAKQAORNTQYSQE 241  
 Db 184 YEQGIQSTTVRHLLVFLYFLPMMGKIKLDTYKELQELNEQTLRYDMEYRSHIE 243  
 Qy 242 SLYKDTSFRHDYLNLTSLRGLENKDLASIEKIYHOLEKTGHQDTRYNIGHANT 301  
 Db 244 ELYKEIRSFRHDYLNLTSLRGLENKDLASIEKIYHOLEKTGHQDTRYNIGHANT 303  
 Qy 302 QNDAVKGILSAKLEAQNKKIAVNVESSKIQPEMELDFTILSTLICDNAEAFSL 361  
 Db 304 RDRALKSLLAGKFKAREKNIVFVNVEPEIQLVEGMSLDFLFTIVSLCDNAEVSARS 363  
 Qy 362 NPEIOLAFKKNGSVIPIONSPKEKODVKIFKENYSWKGNSRGIGLAKVHILEYP 421  
 Db 364 OHVYSTAFLKNGAQETFIENIKEGIDISETSFGASSKGBERGVGLYTVMKIVESHP 423  
 Qy 422 KTSQTSHHHLFKOLII 440  
 Db 424 NTNLNTTCQNOVFRQVLTV 442

RESULT 2  
 US-08-879-941-2 Application US/08879941  
 ; Sequence 2, Application US/08879941  
 ; Patent No. 6268172  
 GENERAL INFORMATION:  
 NUMBER OF SEQUENCES: 6  
 TITLE OF INVENTION: NOVEL HISTIDINE KINASE  
 ADDRESS: Dechert Price & Rhoads  
 STREET: 997 Lenox Drive, Building 3, Suite 210  
 STATE: NJ  
 CITY: Lawrenceville  
 COUNTRY: USA  
 ZIP: 08543  
 COMPUTER READABLE FORM:  
 COMPUTER: IBM Compatible  
 MEDIUM TYPE: Diskette  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/879,941  
 FILING DATE: 08/08/2002  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Bloom, Allen  
 REGISTRATION NUMBER: 29,135  
 REFERENCE/DOCKET NUMBER: Gm10021  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 609-520-3214  
 TELEFAX: 609-520-3259  
 TELE:  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 446 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-879-941-2

---

Query Match 34.6%; Score 752; DB 4; Length 446;  
 Best Local Similarity 39.2%; Pred. No. 8 9e-58;  
 Matches 172; Conservative 86; Mismatches 169; Index 12; Gaps 5;

Db 66 PFLFGILSFLLRLGPKKLILYGFPMILVEYRGYSYEVPLQGIVDGNPFL 125  
 Qy 122 TYMEFAGIALSYLFSVENNDIGRKLDSLTKMKVKKRLLPMMTMLYLLQVWIE 181  
 Db 6 ILYALVINGLKIVIFFKUNGIGLEFDRIFKAFLKFLGLITTEQFLAVSKYLSFIE 65  
 Qy 65 PYFLTAISIYURQNSLNLNFFYGLPVAASSDERRATIFTDGT-QGIV-MGSII 121  
 Db 66 PFLFGILSFLLRLGPKKLILYGFPMILVEYRGYSYEVPLQGIVDGNPFL 125  
 Qy 122 TYMEFAGIALSYLFSVENNDIGRKLDSLTKMKVKKRLLPMMTMLYLLQVWIE 181  
 Db 126 LIMIFVCIFL-VFLKWDYDTRREFLDTGFOQLSUTKINWAMGAYLWMSYLE 183  
 Qy 182 SYNVIPTLKFRKVVTYLILFLILTSLSQYTKQVKVNEIMAKQAORNTQYSQE 241  
 Db 184 YEQGIQSTTVRHLLVFLYFLPMMGKIKLDTYKELQELNEQTLRYDMEYRSHIE 243  
 Qy 242 SLYKDTSFRHDYLNLTSLRGLENKDLASIEKIYHOLEKTGHQDTRYNIGHANT 301  
 Db 244 ELYKEIRSFRHDYLNLTSLRGLENKDLASIEKIYHOLEKTGHQDTRYNIGHANT 303  
 Qy 302 QNDAVKGILSAKLEAQNKKIAVNVESSKIQPEMELDFTILSTLICDNAEAFSL 361  
 Db 304 RDRALKSLLAGKFKAREKNIVFVNVEPEIQLVEGMSLDFLFTIVSLCDNAEVSARS 363  
 Qy 362 NPEIOLAFKKNGSVIPIONSPKEKODVKIFKENYSWKGNSRGIGLAKVHILEYP 421  
 Db 364 OHVYSTAFLKNGAQETFIENIKEGIDISETSFGASSKGBERGVGLYTVMKIVESHP 423  
 Qy 422 KTSQTSHHHLFKOLII 440  
 Db 424 NTNLNTTCQNOVFRQVLTV 442

RESULT 3  
 US-09-747-116-2  
 ; Sequence 2, Application US/09747116  
 ; Patent No. 6348340  
 GENERAL INFORMATION:  
 APPLICANT: Wallis, Nicola  
 TITLE OF INVENTION: NOVEL HISTIDINE KINASE  
 NUMBER OF SEQUENCES: 6  
 ADDRESS: Dechert Price & Rhoads  
 STREET: 997 Lenox Drive, Building 3, Suite 210  
 CITY: Lawrenceville  
 STATE: NJ  
 COUNTRY: USA  
 ZIP: 08543  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/747,115  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Bloom, Allen  
 REGISTRATION NUMBER: 29,135  
 REFERENCE/DOCKET NUMBER: Gm10021  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 609-520-3214  
 TELEFAX: 609-520-3259  
 TELE:  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 446 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-879-941-2



TELEFAX: 610-270-5090  
 TELEK:  
 INFORMATION FOR SEQ ID NO: 326:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 105 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: No. 6340328E  
 US-08-858-207A-326

---

Query Match 8.8%; Score 192; DB 4; Length 106;  
 Best Local Similarity 41.8%; Pred. No. 9.1e-10; Indels 151; Gaps 21;  
 Matches 38; Conservative 21; Mismatches 32; Indels 0; Gaps 0;

Qy 24 LFSKVNTVLSKELTFSISNLFMIAVTVNVNLFPAPLFTIASIYL-NRQNSL 82  
 Db 6 LFTKFLIVTFS---IFSVLVICLHLAI-----YFLFPSTYLSHRQETIG 46  
 Qy 83 LNTYGLPVAASDLFRRAITEFIELD----GTOGIVMSIITVMIEFAGIALSYLF 136  
 Db 47 OKATAIAQSLEGD---RQEQLVDLISQTSDFIKTVGE----MTE----- 87

Qy 137 LSVENVDIGRLKSL-----TKMKVKKLIPKNNI-TMILYVLL-----QVLY 178  
 Db 88 -----DKLEVVKPSLPLDTDROTSFLIEERKVQTDGGTMILQFLASMDLOKEAEQISL 141

Qy 179 VIESYNVPIKFLSLLVAYARTIVAPILETKRVTR---MMDLDSQVRLVSKD 231  
 Db 142 OFLPYTLTFLSLLVAYARTIVAPILETKRVTR---MMDLDSQVRLVSKD 195

Qy 232 NTOYSQDIESLYKD-----IRSFHDYUNLTSLRIGE 266  
 Db 196 EIGNIKEQDNEOYHLLVIAUDHEKNEATLQLEKMKVFLRGASHELKTPLASLKLIE 255

Qy 267 N-----K-----LASIEKYHQIIE-----KGQIQLQTRYNIGHALNIQD 306  
 Db 256 NMRENIGRYKDRDQYLGVALGAVDELNHHVQILSLSVQELRDRETTIDLOMTQNLK 315

Qy 307 KGJUSAKTLEQ-----NRKAVNVEVSSIQIEMELLDFITLSTICNAEAES 360  
 Db 316 DYALIAKERELQDNLSTHQAYLNPSV-----MKL---ILSNLTSNAIKHSY-- 360

Qy 361 LNPEIQLAFFKNGSIVPIQNSTKEQDLSKIFKENYSTKGSNRGIGLAKVNHILEH 419  
 Db 361 --PGGLVRIGEREGEL-FIENSCEBQEKLAQSFDNASRKVKVGSGMLFVVKSLLH 416

---

RESULT 6

US-09-081-689-2

Sequence 2, Application US/09081689  
 Patent No. 6168992  
 GENERAL INFORMATION:  
 APPLICANT: Wallis, Nicola G.  
 APPLICANT: Galacain, Madgalena  
 APPLICANT: Throup, John  
 APPLICANT: Biswas, Sanjoy  
 TITLE OF INVENTION: Histidine Kinase  
 NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Dechert, Price & Rhoads  
 STREET: 4000 Bell Atlantic Tower, 1717 Arch Stree  
 CITY: Philadelphia  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19103-2793

COMPUTER READABLE FORM:  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: Windows 95 Version 2.0b  
 SOFTWARE: FASTSEQ FOR Windows Version 2.0b

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09-081,689  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/046,347  
 FILING DATE: 30-MAY-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Falk, Stephen T  
 REGISTRATION NUMBER: 36,795  
 REFERENCE/DOCKET NUMBER: GM10009  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-994-2488  
 TELEX: 215-994-2222

INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 442 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-09-081-689-2

---

Query Match 6.7%; Score 145.5; DB 4; Length 442;

APPLICATION NUMBER: 60/043, 489  
FILED DATE: 10-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Falk, Stephen T  
REGISTRATION NUMBER: GM50025  
REFERENCE/DOCKET NUMBER: 36-795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2488  
TELEX: 215-994-2222  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 363 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

---

Query Match 6.5%; Score 141.5; DB 3; Length 363;  
Best Local Similarity 19.7%; Pred. No. 0.00013;  
Matches 84; Conservative 82; Mismatches 126; Indels 135; Gaps 20;

QY 36 KELTFISNLML-AVTMWNFLPAAPIKFLVNLNSLNLNFGSLPPAS 94  
Db 5 KDTSTAESSSILFLIEPIAGIFPVNEVYGPKMLYLIVIPR----SLS---YLILTVN 55

QY 95 SDLFRATIFFILDGTGOGIVMGSSIITTYMEFAGIALSTFLSVNFDIGRLKDSLTM 154  
Db 56 NRL--NTMFLYI----LIIHFLCYFVSYHPMSLFVPEAIPF-TENNNVKT 107

QY 155 KVKKRLIPMMTMLYYLLQVLYVYESYNVPTLKRKFWVVIVVLFILLISFLSQV 214  
Db 108 ATNLFLTMICITIYTLYNNFVA-----MMVYVVISLMDUNFKMK 153

QY 215 RQKVONETMAKEAQIRNTTQSQQESLYKDORSFRHDYVNLTSLRUGENKLASTE 274  
Db 154 NREQKEI-AEKNRHINL----- 171

QY 275 KIVHOILEKGHQHQLDTRYNHLANTONDAVKGTSAKITEAQNKI---AVNV--- 326  
Db 172 -IAEQRHIGQDHT--LGIV--FASLISLSELAJKLTDYEVKAELINKSR 225

QY 327 -----EVSKIQOLPE-MELDFITLISLIDNATEAFAFSLNPIBOLAFFKRNSTV 378  
Db 226 ESLNKVREIDDVKLPSFIED-SIRVVLKADDFTE--NKEAQVLSPKOSMV 281

QY 379 IIONSTKEQIDVKISKIFKENYSFGNSNRGIGAKVNTHLEIYPTK----SIQTNHHHF 434  
Db 282 MI---TRE-----AINNVIKHANASKVHGRLKTVNNH--- 310

QY 435 KOLIQT 441  
Db 311 KILLMIE 317

---

RESULT 8  
US-09-046-086-2

SEQUENCE 2: Application US/09046086  
PATENT NO. 6127147  
GENERAL INFORMATION:  
APPLICANT: Wallis, Nicola G.  
TITLE OF INVENTION: Novel Compounds  
NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:  
ADDRESSSEE: Deehert, Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stree  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2713  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/046,086  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:

---

RESULT 9  
US-09-299-378-4  
SEQUENCE 4: Application US/09299378  
PATENT NO. 634555  
GENERAL INFORMATION:  
APPLICANT: Higgins, Lauren S.  
APPLICANT: Dalton, Michael A.  
APPLICANT: Kong, Huimin  
TITLE OF INVENTION: Method For Cloning And Producing The SWAI Restriction File Reference: SWAI  
FILE REFERENCE: SWAI  
CURRENT APPLICATION NUMBER: US/09/299,378  
CURRENT FILING DATE: 1999-04-27  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 618

; TYPE: PRT  
 ; ORGANISM: Staphylococcus warneri  
 ; US-09-299-378-4  
 Query Match 5.4%; Score 117; DB 4; Length 618;  
 Best Local Similarity 21.4%; Pred. No. 0.038; Gaps 27;  
 Matches 116; Conservative 76; Mismatches 187; Indels 164; Gaps 27;  
 Db 92 IENTLILTKKKVLDRACTGTFSTAFIHEFNROISKNSIVDF-INNYLNTIDSNEMINF 150  
 Qy 8 LSGNLITYLTVL-----FLULFSKVSNVNLSSKLTLISNSLI-----MIAV 52  
 Db 53 TWVN-----VNLYPAEP-LYFI-----ALSIYLNQNSLSN---- 84  
 Db 151 TKINTLTMYYLYNDISIDFKVKNIYAIDFVYQEKHKEFNLNFYNSONGNFINDNFEN 210  
 Qy 85 --IFYCLLPASSDIF-RRAI-----TEFLDGTQ 112  
 Db 211 FDIVIGNPPVS--LYGRRAINKEDKROFLRINMYDFIPKNVNGKKNYTHFENGQL 268  
 Qy 113 IWMGSSI-----ITTYMIEFAGIALSYFLSVN-----VDIGRKDSL 151  
 Db 269 LKKGNLITFWDITLESRESIJKYLETALKOLDNLNKNSEDVYSGQILSLKNA 328  
 Qy 152 TRMKV-----KKRLLPMMI-----YVLL-----QVLYVES-YVNPITLK--PRKF 194  
 Db 329 NKEAIVSIKOMQWNNTIQINQDINWHDKFYRFKNSDKKINSLEKVKNSDELQYFPKK 388  
 Qy 195 VVIVVYLFLFLITLISQYKQYKVNEMIAQKEQIRITQSQQES-----LYK 245  
 Db 389 EIRTSTMLNNMSESSFVVDY-KPETYDFHVPAYKKG-AKNLSPFPQNMHSNHYFTDIALQK 446  
 Qy 246 DIRSERHDY-----NLTTSRGLGENKDASIEKY----HOILEKGTHQLODTRYNIGH 297  
 Db 447 KIDSHEELLIKKGKRNKRIGLGLEVKNPKLFIROSANKNLIATDGKMSRANSLYI 506  
 Qy 298 LANIQNDAVKGILSAKILEAQNKKAFLAVNVVEVSSKIQPEMELDFITLSTLDDNAA 357  
 Db 507 LSKATND-TKDINMKITCAQLNS-----ELTTFIAL---TNTRIRK 544  
 Qy 358 FESLNEIQLAFLFKRNGSIVFIQNSTKQKQYKSENKENTSTKGSURGIGLAKNHIL 417  
 Db 545 ARBKQPKQLSDLK---TIPLCFNEBEINSKLL---IFEAENAKKNNBLESSELIKNOIT 597  
 Qy 418 EHY 420  
 Db 598 YKY 600  
 RESULT 10  
 US-08-963-001-2  
 Sequence 2, Application US/08963901  
 ; Patent No. 6270992  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wallis, Nichola G.  
 ; APPLICANT: Traini, Christopher M.  
 ; APPLICANT: Kosmatka, Anna L.  
 ; APPLICANT: Shilling, Lisa K.  
 ; APPLICANT: Warren, Richard L.  
 ; TITLE OF INVENTION: NOVEL HISTIDINE KINASE  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Bechtel Price & Rhoads  
 ; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: US  
 ; ZIP: 19103  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 RESULT 11  
 US-08-588-985-2  
 ; Sequence 2, Application US/08588985  
 ; Patent No. 577094  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Michiuki MATSUDA et al.  
 ; TITLE OF INVENTION: cDNA OF DOCK180 GENE AND DOCK180 PROTEIN  
 SOFTWARE: FastSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/963,901  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Dickinson, Todd O.  
 REGISTRATION NUMBER: GM10058  
 REFERENCE/DOCKET NUMBER: GM10058  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-994-2232  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 457 amino acids  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 ; US-08-963-901-2  
 Query Match 5.2%; Score 114; DB 4; Length 457;  
 Best Local Similarity 19.6%; Pred. No. 0.045; Gaps 24;  
 Matches 96; Conservative 76; Mismatches 167; Indels 152; Gaps 24;  
 Db 9 IAIKSITVILSALSALISFLVITNYHYNIKASNDAKIMTKEAROYEQSAPKPHIQYFK 68  
 Qy 38 LTLESISNLFNLIMAVTMVNVLFY-----F 73  
 Db 69 HLGMQNYQTMVDOOKHKTFG-EPEFRDTISONAINNLNNDKDHGIKDKPFALFVG 127  
 Qy 105 F---TLDGTQGIVMGSSIITTYMIEFAGIALSYFLSVNHYDGR-LKDSLTKMKVKRLI 161  
 Db 128 FDNYTDNTVGI-----NFKTKDGSIAV-----FMRPDIGETSEFRFLAV----- 168  
 Qy 162 PMNITMLLIVLILQVIVVIESNVPITKFRKFWVIVLFLILISELSQYTKVONE 221  
 Db 169 ---LMLLFLFISIS-LVIASTYTSITRVPKK-----LATERLIDGDFEPTIKQTKE 218  
 Qy 222 I-----MAQEAQRNTIQSQO-----TESLYKI---RFRHDYL 255  
 Db 219 IGTLOQHENKMRBLSLQYDQMRHOFVONVSHETKPLTHIHLUSLBQTSKDLRQYI 278  
 Qy 256 N---ILTSLRGLGENKDASIEKYHQILEKGTHQLODTRYNIGHANIQNDAVKGILSA 312  
 Db 279 NDIVYHTQSLGUTELULLSLENDHQ-----HLFDDKTOVOL-----IKDIIRH 325  
 Qy 313 KILEAQNKKAFLAVNVVEVSSKIQPEMELDFI-----TILSILCONAIAFESLNPET 365  
 Db 326 EQFAADEKSLII-----LADLESINFQNLQRIHQALQNSLNLNAIK---YDVGGA 374  
 Qy 366 QLAFFKKNSIVFIQN-----STKERQDVKSTKRENY--STKGSNRQIGLAKVNHLE-HY 420  
 Db 375 DIALOSHNNIIFIISNDGSPSPOAE-ARLFERYKVSKDNNSGLATKTSIELH 433  
 Qy 421 PKTSQTSHHH 431  
 Db 434 GTIQFTQSNEY 444

NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Wenderoth, Lind & Ponack  
 STREET: 805 Fifteenth Street, N.W., #700  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: U.S.A.  
 ZIP: 20005  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 5.25 inch, 500 kb  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: Wordperfect 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/588,985  
 FILING DATE: January 19, 1996  
 CLASSIFICATION: 536  
 PRIORITY APPLICATION DATA:  
 REFERENCE/DOCKET NUMBER:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Warren M. Cheek, Jr.  
 REGISTRATION NUMBER: 33,367  
 REFERENCE/DOCKET NUMBER:  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-371-8850  
 TELEFAX:  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1865 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL:  
 ANTI-SENSE:  
 FRAGMENT TYPE:  
 ORIGINAL SOURCE: spleen cell of homo sapiens  
 US-08-588-985-2

Query Match 5.2%; Score 113; DB 1; Length 1865;  
 Best Local Similarity 20.1%; Pred. No. 0.39; Matches 95; Conservative 90; Mismatches 208; Indels 80; Gaps 21;

QY 21 LIFLFSKVSNVTLKELTFLSISNPLIMIATVMN-VNLIFYPAEPYLIALSIYN-RQ 78  
 Db 772 LIQLFRSINDMSSMSDQTV-RVGAAALKYLPITIVNDVKLVPDFPKLSKMTTEFLINPM 830

QY 79 NSLSNIFYGSILPVASSDLFRRAIFFILDGTGIGVAGSSITYM---IEFGIALS 133  
 Db 831 GILTIIOKLYCLIEVHSDFLTQHDCREIL---LPMMTDQKYLHLEQDEACCOLIS 885

QY 134 YLFSLFVNVIDGRKLSDTLMKVKRKLIPMTML-----LYLILQVIES 182  
 Db 886 HILEVLVYRKDGPTQRHV-QIMEKLLRTVNRFTVISMGRDSELIGNFVACMTAIIROMED 944

QY 183 YNVIPI TLK-FRKVVWYLVLFILFLISFLSFTSYTKQ-----KVQEIMAKQEAOQNTRQ 235  
 Db 1003 YADMILKKFLDQANFELQLWNWYFHIAVAFLQESLOLENFSAKRAKILNKYGDMMROI 1062

QY 285 GHQDQTYRNG-HLANTQNDVKGTLISKYL-EAQNKKAVN-----EVSSIQ 333  
 Db 1063 YSQQIESLYKDTSFR---HQLNTI---LTSRLGJENKLASIEKI---HQIILEKT 284  
 Qy 79 NSLSNIFYGSILPVASSDLFRRAIFFILDGTGIGVAGSSITYM---IEFGIALS 133  
 Db 831 GILTIIOKLYCLIEVHSDFLTQHDCREIL---LPMMTDQKYLHLEQDEACCOLIS 885

Qy 134 YLFSLFVNVIDGRKLSDTLMKVKRKLIPMTML-----LYLILQVIES 182  
 Db 886 HILEVLVYRKDGPTQRHV-QIMEKLLRTVNRFTVISMGRDSELIGNFVACMTAIIROMED 944

Qy 183 YNVIPI TLK-FRKVVWYLVLFILFLISFLSFTSYTKQ-----KVQEIMAKQEAOQNTRQ 235  
 Db 1123 MFENEI--ITKL---DHEVGGRGDEQYKVLFDRILLERHKHLAKTGTGFVKLVV 1175

RESULT 12  
 US-08-971-988-2  
 Sequence 2, Application US/08971988  
 Patent No. 5786461  
 GENERAL INFORMATION:  
 APPLICANT: Michiyuki MATSUDA et al.  
 TITLE OF INVENTION: cDNA OF DOCK180 GENE AND DOCK180 PROTEIN  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Wenderoth, Lind & Ponack  
 STREET: 805 Fifteenth Street, N.W., #700  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: U.S.A.  
 ZIP: 20005  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 5.25 inch, 500 kb  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: Wordperfect 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/971,988  
 FILING DATE: 17-NOV-1997  
 CLASSIFICATION: 536  
 PRIORITY APPLICATION DATA:  
 REFERENCE/DOCKET NUMBER: 08/588,985  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Warren M. Cheek, Jr.  
 REGISTRATION NUMBER: 33,367  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-371-8850  
 TELEFAX:  
 TELEX:

INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1865 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL:  
 ANTI-SENSE:  
 FRAGMENT TYPE:  
 ORIGINAL SOURCE: spleen cell of homo sapiens  
 US-08-971-988-2

Query Match 5.2%; Score 113; DB 1; Length 1865;  
 Best Local Similarity 20.1%; Pred. No. 0.39; Matches 95; Conservative 90; Mismatches 208; Indels 80; Gaps 21;

QY 21 LIFLFSKVSNVTLKELTFLSISNPLIMIATVMN-VNLIFYPAEPYLIALSIYN-RQ 78  
 Db 772 LIQLFRSINDMSSMSDQTV-RVGAAALKYLPITIVNDVKLVPDFPKLSKMTTEFLINPM 830

QY 79 NSLSNIFYGSILPVASSDLFRRAIFFILDGTGIGVAGSSITYM---IEFGIALS 133  
 Db 831 GILTIIOKLYCLIEVHSDFLTQHDCREIL---LPMMTDQKYLHLEQDEACCOLIS 885

QY 134 YLFSLFVNVIDGRKLSDTLMKVKRKLIPMTML-----LYLILQVIES 182  
 Db 886 HILEVLVYRKDGPTQRHV-QIMEKLLRTVNRFTVISMGRDSELIGNFVACMTAIIROMED 944

QY 183 YNVIPI TLK-FRKVVWYLVLFILFLISFLSFTSYTKQ-----KVQEIMAKQEAOQNTRQ 235  
 Db 1123 MFENEI--ITKL---DHEVGGRGDEQYKVLFDRILLERHKHLAKTGTGFVKLVV 1175

QY 945 YHYALIKFGKMRDWDLFMLFIMPKNLIGKNNYFPDWYMN - MVOKVYFLRAINO 1002

QY 236 YSOQTESLYKKDIRSFR---HDYNI ---LTSRLGTEKNDIASTEKI ---YHOILEKT 284

Db 1003 YADMINKEDDANFELQIWNWTFHAYAFTQESOLESSAKRAKILNYKGDRHQI 1062

QY 285 GHQLODTRYNG-HLANTONDAGVILSALKL-EAQNKKAVN-----EVSKIQ 333

Db 1063 GFEIRDMWNLNQHKIKFPEMFQPILEMLTIPETELRKATIPIFDMQMCEFHSRSFO 1122

QY 334 LPLEMILDEFTISILCDMIAIEA----FESINPELOL-----AFPKKNSIVTION 382

Db 1123 MFENEI--ITKL---DHEVEGGRGDFOYKVLFKILLERCRKHKYLAKTGETFWKLW 1175

QY 383 STKEKQDVKSFKENYSKGNSRNGIGLAKKVNILLEHPKTSLOTSNNHHFK 435

Db 1176 RLMERLIDDTTMIHD-----ENKNRMSTCVANLNFYKEIREEMVIRYLIK 1222

RESULT 13

US-08-462-949-2

; Sequence 2, Application US/08462949

; Patent No. 5006022

GENERAL INFORMATION:

APPLICANT: Rasmussen, Beth Ann

TITLE OF INVENTION: Cloning and Identification of a Two Component Signal Transducing Regulatory System from Title of Invention: Bacteroides Fragilis

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESS:

ADDRESSEEE: Darby & Darby P.C.

STREET: 805 Third Avenue

CITY: New York

STATE: New York

Country: U.S.A.

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/462, 949

FILING DATE: 08/02/2001

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/023, 764

ATTORNEY/AGENT INFORMATION:

NAME: Robinson, Joseph R.

REGISTRATION NUMBER: 33,448

REFERENCE/DOCKET NUMBER: 0646/1B024-US1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-527-7700

TELEFAX: 201-753-6237

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 519 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI SENSE: NO

US-08-462-949-2

Query Match 5.2%; Score 112.5; DB 1; Length 519;

Best Local Similarity 21.8%; Pred. No. 0.073;

Matches 73; Conservative 64; Mismatches 127; Indels 71; Gaps

QY 129 GIALSYLFSLFSVEND-----IGRKLDSL-----KMKVKRLLPMTM 167

RESULT 14  
 US-08-023-764B-2  
 ; Sequence 2, Application US/08023764B  
 ; Patent No. 5679540  
 ; GENERAL INFORMATION:  
 APPLICANT: Rasmussen, Beth Ann  
 TITLE OF INVENTION: Cloning and Identification of a Two Component Signal Transducing Regulatory System from Title of Invention: Bacteroides Fragilis  
 NUMBER OF SEQUENCES: 39  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: American Cyanamid Company  
 STREET: One Cyanamid Plaza  
 CITY: Wayne  
 STATE: New Jersey  
 COUNTRY: United States  
 ZIP: 07470-8426  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: IBM PC compatible  
 SOFTWARE: PC-DOS/MS-DOS  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/023,764B  
 FILING DATE: 26-FEB-1993  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Baird, Elizabeth M.  
 REGISTRATION NUMBER: 31,008  
 REFERENCE/DOCKET NUMBER: 31,658-00  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (201)831-3246  
 TELEX/FAX: (201)831-3305  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 519 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-023-764B-2  
 QY 193 GIELLYH-SVVIDKGREVRCSDYEEGGSDESYTQLFONDPPARMSTVKVHFPGKKD-250  
 Db 168 LLYYLQIQLIVIESYN-VPTLKFKEVVVYLILELILISFLSOTYKVKVQ-NEIMA 224  
 ; :  
 Db 251 -----YIFVSFSEMPISMF-TFVLITPFTIYIVYFROKKLNMDNFIINWTH 299  
 QY 225 QKEAQIRNTIQYSOQE--SLYKIDRSFRHDYLNLNTSLRGLENKDLASEKIVHOLE 282  
 ; :  
 Db 300 EFKTPISTISSLAAQMLKDPAVGKSPOMFOH-----ISGVINDETKRLFQVEKVLO 350  
 QY 283 KTHGHOQDFTYRIGHNLNTQNDAVKGILS--AKILEQNKIAVNNEVSSK-IQPEMEL 339  
 ; :  
 Db 351 MSMFDRQKATLKMREDL-ANELLSGVNTALKVERVNGKILTSNLNEATVIFADEMH 408  
 QY 340 LDFITLISICDNIAEAFFESLNEPIQIAF---FKKNSIVFIO-NSTKEKQDVKI 394  
 ; :  
 Db 409 TN--VIFNMNDNAVKYK--KPEEDIIVLVDTVTWNPFGKLMSIODNGIGGIKENLK 461  
 QY 395 FKENYSTKGSSNR----GILGLAKVNHILEHYKRT 423  
 ; :  
 Db 462 FDKEVRYVHGNLHDVKGFGLGLAVVKKIQDHKG 496

Db 193 GIELLYHF-SVVDKDGREVRCSDYEEGSEDSITQPLFONDPPAKMSIVKVHFPKKD- 250  
 Qy 168 LLYVLLIQVLYVIESN-VIPTLKERFKVIVYLFLFLISLFSQYKOKVO--NEIMA 224  
 Db 251 -YIFDSVSPMFPMIF-TFVLITIFTYIVFRQLKTEMKNDFINNMTH 299  
 Qy 225 OKEAQIRNITOVSQOLE--SIYKDRSRFDYLNITSRLGLENKDASIEKIHOLE 282  
 Db 300 EFKTPISTISLAQMLKDPAVGKSPOMFH----ISGVINDETRKLRFQVEKVHQ 350

Qy 283 KIGHOLQDTRYIGHLANIQDAVGILS-AKLEAQNKAVNVESK-TQLPENEL 339  
 Db 351 MSMFDROKATLMKEDL--ANELISCVINTFLAKVYRNLNTSLEAMNPVFADEMHI 408  
 Qy 340 LDPITISLICNAEAFESLNPEQALF---PKKNSIVFIQ-NSTKEQIDISKI 394  
 Db 409 TN---VIFNLMDNAVVKY---KPEEDLVDVYTWNNEPGKLMISIQDNGIGIRKENLKVV 461  
 Qy 395 FKENYISTKGNSR----GIGLAKVHILEHYPKT 423  
 Db 462 FDKFYRHTGNLHDVKFGUGLGLAYVKKILODHGT 496

RESULT 15  
 US-09-081-686-2 Application US/09081686  
 ; Sequence 2, Application US/09081686  
 ; Patent No. 6162019

; GENERAL INFORMATION:  
 ; APPLICANT: Wallis, Nicola G.  
 ; APPLICANT: Shilling, Lisa K.  
 ; APPLICANT: Wang, Min  
 ; APPLICANT: Jaworski, Deborah D.  
 ; APPLICANT: Ingram, Karen A.  
 ; APPLICANT: Yigong Ge, James  
 ; APPLICANT: Holmes, David J.  
 ; APPLICANT: Zalaiajn, Magdalena  
 ; APPLICANT: Throp, John  
 ; APPLICANT: Biswas, Sanjoy  
 TITLE OF INVENTION: Histidine Kinase  
 NUMBER OF SEQUENCES: 9  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Dechert, Price & Rhoads  
 STREET: 4000 Bell Atlantic Tower, 1717 Arch Street  
 CITY: Philadelphia  
 STATE: PA  
 ZIP: 19103-2793

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: Windows<sup>95</sup>  
 SOFTWARE: FastSeq for Windows Version 2.0b  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/081,686  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/048,339  
 FILING DATE: 30-MAY-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Falk, Stephen T  
 REGISTRATION NUMBER: 36 795  
 REFERENCE/DOCKET NUMBER: GM10006  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-994-2488  
 TELEX: 215-994-2222

INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 446 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single

; TOPOLOGY: linear  
 ; US-09-081-686-2

Query Match 5, 0%; Score 108; DB 4; Length 446;  
 Best Local Similarity 18.3%; Pred. No. 0.15; Index 218; Gaps 23;  
 Matches 96; Conservative 78; Mismatches 133; Indels 218; Gaps 23;

Db 6 KLUTINSFLASSFAILGGVGIWTAIYLFFPYTOLP--CEKENSARVETEVYKTKS 52  
 Qy 96 DLFRRAIIFFLDGQTQGIVMGSSITTYMIEFAGIALSY--LFLS----- 138  
 Db 63 D-----ELPSLQSYOSRSLTSAHLRDRIVKRLP 93

Qy 139 VFNVIDT--GRLKDSLTKMKVKKRLLPMNI-TMILYUILLQVIVXIESYNVITPLKRFV 195  
 Db 94 VHDLUDIKDGKLSNVIV---MLDMSYSTADGKQTVQFHGVQV---KEAKNI 140  
 Qy 196 VIVYL-ILFLFLISF--LSQYTKOKVONET-----MAKENQIR-----NITQ 235  
 Db 141 LLYVLPYTEVTLIAFSVIVFYVTKRLLNPLVISEVTSKMQDLDNIRDESDRKEVGE 200

Qy 236 YSQQIESLYKD-----IRSFHDYLNITSRLGJEN-- 267  
 Db 201 VGKQINGMFBHLKVTHELESRNREQIVKLNQKVSVFVGASHELKPLASLRILNMHQ 260

Qy 268 -----KOLASLEKIVH--QILEKIGHQ-----LOD 290  
 Db 261 NIGDYKDHPKYIKAISINKIDOMSHLLEEVELESSKFOEWTECRETLVKPVLYDILSRYQE 320

Qy 291 TRYNG-HLANTIONDAVGILSANSAMLEQNKAVNVESVSRQIOLPEMELIDFTILSIL 349  
 Db 321 LAHSIGVTENQUDTATRVMSLRALD-----KVLTNL 353

Qy 350 CDNA:AAFLSINPEQIALLFKKNSIVFIQN--STKE---KQIDYSKIFRENY-- 399  
 Db 354 ISNATK-----YSDKNGRVLISEQDYLSTCAPLSQDEELHLDFTYHS 400

Qy 400 ---STKGNSNGIGLAKVHILEHYP-----KSLQT 427  
 Db 401 QIVTDKDESSGIGLVIVSNLESQYDMRDYLPYEHGMFHKISLT 445

Search completed: July 30, 2002, 15:14:44  
 Job time: 520 sec

Wed Jul 31 08:30:56 2002

Page 10

us-09-833-017-4.rai

Copyright (c) 1993 - 2000 CompuGen Ltd.  
GenCore version 4.5

OM protein - protein search, using sw model  
Run on: July 30, 2002, 15:15:18 ; Search time 29.21 seconds  
Perfect score: US-09-833-017-4 (without alignments)  
Sequence: 2173 1 MNEALMILSMLGLLTYLTVLF. .... KTSIQLTSHHHFKQLLIK 441  
Scoring table: BLOSUM62  
Gapext 10.0 , Gapext 0.5  
Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000!  
Post-processing: Minimum Match 0%!  
Maximum Match 100%  
Listing First 45 summaries  
Database : PIR\_71:  
1: pir1:\*
2: pir2:\*
3: pir3:\*
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query	Length	DB ID	Description
1	755	34.7	446	H97329	histidine kinase (sensor histidine kinase)
2	749	34.5	446	D95061	histidine kinase-1 (probable sensor histidine protein kinase)
3	455	20.9	447	F96509	sensor histidine k
4	427.5	19.7	441	B93261	sensor histidine k
5	423.5	19.5	441	G98126	sensor histidine k
6	418.5	19.3	431	A01081	sensor histidine k
7	416	19.1	431	AD1438	sensor histidine k
8	379	17.4	423	A95480	agrc protein - Sta
9	372	17.1	371	D89995	accessory gene reg
10	362	16.7	428	G56219	protein histidine
11	348	16.0	448	S57909	probable histidine membrane associate
12	223.5	10.3	452	B97195	two-component sens
13	184.5	8.5	433	A01533	two-component sens
14	174	8.0	542	E69600	two-component sens
15	171.5	7.9	532	E83569	two-component sens
16	163.5	7.5	657	D81508	probable integral
17	155.5	7.2	440	C55221	virs protein - Clo
18	154	7.1	533	I17009	two-component sens
19	151.5	7.0	921	G71705	alkaline phosphata
20	151	6.9	543	D65222	hypothetical 60.6
21	151	6.9	543	C91267	two-component sens
22	151	6.9	543	H86107	probable 2-compone
23	150.5	6.9	543	AH1033	two-component sens
24	148.5	6.8	473	G97085	sensory transduct
25	146	6.7	1447	F82909	hypothetical prote
26	145.5	6.7	442	C95070	sensor histidine k
27	145.5	6.7	442	A97938	vines, histidine ki
28	143	6.6	538	G84129	two-component sens
29	6.6	885	2	D90000	sensor protein Kap

#### SUMMARIES

```

Total number of hits satisfying chosen parameter: 1
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%!
Maximum Match 100%
Listing first 45 summaries
Database :
          PIR_71:*
1:  pir1:*
2:  pir2:*
3:  pir3:*
4:  pir4:*

```

pred. No. is the number of results predicted to have a score greater than or equal to the score of the best hit and is derived by analysis of the total score.

ARGUMENTS

Copyright (c) 1993 - 2000 Compugen Ltd.  
 GenCore version 4.5  
 Run on: July 30, 2002, 15:15:18 ; Search time 29.21 seconds  
 OM protein - protein search, using sw model  
 Sequence: 1 MNEALMLILSNGLILTYLTVLF . . . . . KTSIQLTSNHHHLFKQLLTIK 441  
 Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

30	142	6.5	824	2	A867133	glycosyl transfera
31	138.5	6.4	363	2	B89907	hypothetical prote
32	137	6.3	428	2	S32935	kinB protein
33	136.5	6.3	1008	2	F71727	acriflavin resista
34	136	6.3	733	2	C89940	hypothetical prote
35	136	6.3	754	2	G7014	hypothetical prote
36	135.5	6.2	542	2	E90604	hypothetical prote
37	135	6.2	569	2	B70113	hypothetical prote
38	134.5	6.2	553	2	AD0579	sensor kinase Dpb
39	134	6.2	565	2	G64532	methyl accepting c
40	133.5	6.1	538	2	D82180	probable sensor ki
41	131	6.0	564	2	F71975	methyl-accepting c
42	130.5	6.0	714	2	C90100	hypothetical prote
43	130	6.0	590	2	G90127	hypothetical prote
44	129.5	6.0	339	2	C96927	sensory transduct
45	129.5	6.0	532	2	H83993	two-component sens

RESULT 2

D95051 sensor histidine kinase Blch, probable [imported] - *Streptococcus pneumoniae* (strain TIC) 414

C;Species: *Streptococcus pneumoniae*  
C;Accession: F96909  
C;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C;Cross-references: GB:AE005672; PIDN:AAK74685.1; PID:g14972002; GSPDB:GN00164; TIGR:SP4  
C;Genetics:  
A;Gene: SP0527

Query Match 34.5%; Score 749; DB 2; Length 446;  
Best Local Similarity 37.8%; Pred. No. 2.2e-33; Matches 170; Conservatve 96; Mismatches 163; Indels 70; Gaps 10;  
Matches 170; Conservative 89; Mismatches 173; Indels 18; Gaps 4;

QY 1 MNEAMIL---SGLLTLVFLIFLSKNSNVTS ---KEELFSSISNLIMA 51  
Db 1 MNIAWLLYLTVLNGL-----EIVIFFKVDGIDLTERIFKAFLKLALLAFVUMIS 52

QY 52 VTMVNVLNPYPAERLYFALLSIYLNRONSLNSIUNIFYGLPVAASSDIFRATTFILDGT- 110  
Db 53 YIWGVNLSPYFMEPDLYGIGSSELLLRGPKKLFYGLPMVNFYRGYSYVFLG 112

QY 111 QGIVNGSSTTYTMYIEFAGIALSYLTSVFNDIGRKDSLTKMKVKRPLPNMTMLY 170  
Db 113 OGQYDGYSTFGCLCILINFFISLAFLKWLDDFTSLRKELDKAQOKSLTQINHMGY 172

QY 171 YLIQOLVLYVESYNYIPTLKFRKEVIVYLFLILISFLSFLQYTKQVONIMAKAQI 230  
Db 173 YLVMESSLSEFEYEQSIQSRTVRHLVFLYFLPMGVKKLDTLYKEKLYERLEOEALRY 232

QY 231 RNTTOYSQOESLKVLDISRFRRHOTLNTTSRIGENKDLASITEKIHOLEKFTGHOLQD 290  
Db 233 RDMEYRSRHEEELKEVERSFRRHOTLNTTSRIGEEDBMOIKEVYGSVVKRDSSQLQ 292

QY 291 TRYNTGLHANIQNDAVKGILSAKILEAQNKIAVNEVSSKIQOLPEMELDFITISLC 350  
Db 293 NKYDGLGRVNNIRDKRAIKSLLAGKFKLARDKNLNVENVEPEILOVEGMSLDFLTIVSILC 352

QY 351 DNAEAFESLNPETIOLAFFKKNSIVIIONSTEKEOIDSKIFKENYSKGSNRGIGL 410  
Db 353 DNATEASVEASQPHSIAVTKLNGAQTTEIENSIKEBGIDISEIFSEGASSKGEERGVGL 412

QY 411 AKVNHTELPKTSIOTSNNHHFLKQLLIT 440  
Db 413 YTMVIVESHPNTSNTTCQQVFRQVLT 442

RESULT 3

F96909 histidine kinase Atpase [imported] - *Clostridium acetobutylicum* 414

C;Species: *Clostridium acetobutylicum*  
C;Accession: F96909  
C;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-447 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK78065.1; PID:g15022902; GSPDB:GN00168  
A;Experimental source: *Clostridium acetobutylicum* ATCC824  
A;Genetics:  
A;Gene: CAC0080

Query Match 20.9%; Score 455; DB 2; Length 447;  
Best Local Similarity 27.4%; Pred. No. 1.6e-20; Matches 124; Conservative 96; Mismatches 163; Indels 70; Gaps 10;

QY 111 GLITYLTYFLFLFSKSVNTSLKETTSISNFLMLAVTAVNVNVLFYPAEPFLFA 70  
Db 33 GMVITEETISLIFLM-----LGWQQTVM----- 70

QY 121 -TTMIEBAGIALSYLFLSVNVDIGRLKSLTKURKVRKLIPIANTMLYYLQIVY 179  
Db 177 WITFCIEFIGTYVTSRTRKVFDR-----KFRAYNGLKSPIGVVAAGTILMFLV 167

QY 180 IESVNTI-----PTLKRKFVWVYLFLFLISFLSOTYTKQVQNEIMAKEA 228  
Db 168 FYTNVNUVFPAPNNSVWNNTMRKGTVLFSEVAILLIVITRILTRGIIKEME--LKSREN 224

QY 229 QIRNITOQSOTESLYKDIIRSFRHDYLNTLSRLGIEENKLASIEKIYHQILEKTGHQ 288  
Db 225 EFOSLQETNKLEKLHKDRGFRHDYINILSMAGYIQNRDLEGIEREFDKIMPLSKM 284

QY 289 QDTRYNTGLHANQDVAKGTSAKLAEONRKAVNEVSSKLTOLPENELDFTILIS 348  
Db 285 KSNFKIGLQNEVPIKGMSAKIRAOETGIDVYDVAESIKSFNNEIIDSRVIGI 344

QY 349 LCNNTAIEAFESLNPEQLOAFLPKKNSIVIIFTIONSTKEOIDVSKIFKENYSKGSNRGIG 408  
Db 345 LDNNAIEASEKCDRPSKWMKVAVINKDSWMIVINNINNEIPPIKIYKGFSIKGDNGI 404

QY 409 GLAKVNHTELPKTSIOTSNNHHFLKQLLIT 441  
Db 405 GLSNLKDIKGYPNVMLDITVEDNQPKQIDIK 437

RESULT 4

B95261 probable sensor histidine kinase Comp [imported] - *Streptococcus pneumoniae* (strain TIC) 414

C;Species: *Streptococcus pneumoniae*  
C;Accession: B95261  
C;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
A;Reference number: A95000; MUID:21357209; PMID:1463916  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-441 <KUR>  
A;Cross-references: GB:AE005672; PIDN:AAK76283.1; PID:g14973747; GSPDB:GN00164; TIGR:SP4  
A;Experimental source: *strain TIGR4*

A;Gene: SP236

Query Match 19.7%; score 427.5; DB 2; Length 441;  
 Best Local Similarity 25.9%; Pred. No. 7.6e-19; Mismatches 188; Indels 53; Gaps 12;  
 Matches 119; Conservative 99; MisMatches 116; Conservatve 84; Indels 77; Gaps 10;

Qy 12 LTYLTIVL--FLFLS--KVSNTLKLTLTISINFLIMAVM-----VNWN 58  
 Db 3 LUGFGTVVHFLLISHSYRLICKGRINKELVYGAETLTLTIEFSFLYLDKGIE 62  
 Qy 59 LFPAAEPDFYLISIYL---NRONSLUNIFGLPVASDLFRRAIFLFDGQ 112  
 Db 63 RF---LFPFLGLYSYFRWMKQYERDQGLFLSLSLVESTHNFL-SVIFESITGDNF 115  
 Qy 113 IV--MGSSITTYMEFAGIALSYLF--LSVFNDIGRLKDSLTKMKVKKALIPMTI 166  
 Db 116 VQYHPEPFVVYVFLTYFVTLRKLIVYFHLEAYFDKY-----LYPLFLK 160  
 Qy 167 MLLYLLIQVLYVIESVINWPLK FRKKVVVYLFLFLISLQSQTQKQONE--I 222  
 Db 161 VFALLLHIVRS- DMSVTKHNSGSSLSSVTFISLUFFFAMNSHKYOMEKETA 218  
 Qy 223 MAKEAQTRNIQYQSOIESLKDRISFRHDLNLTSRGLTNDIASIETYHOLE 282  
 Db 219 LKKKFKEDKHONYTDENGFLYNETGRHDYAGMLVSMOMADSGNLOEIDRINYEVIL 278  
 Qy 283 KTGHOLODTRYNIGHLANIONDAVKGILSAKILEAQNKIAVNEVSSKIQIPEMELDF 342  
 Db 279 KAHHKLRSKDQYFJDLNEDSALRVAQSIYARNGEVTEVKDPIKPIEFLD 338  
 Qy 343 ITILSILCNAAEFAESLNPFQIOLAFFKKNGSIVFTIONSTKEQDVKIRNEYSTK 402  
 Db 339 VRIMSVLNNNAEGSADSQSKYQMEVAKMPTETVIVIONSKMMTPSGDIFALGSTK 398  
 Qy 403 GSNRGIGLAKVHILEHYPKTSQTSNHHHLFKQLLIK 441  
 Db 399 GRNRGVGLNNVKEELDKYNNILETEMEGSTFRQIIRFK 437

RESULT 5

G98126

histidine protein kinase (EC 2.7.3.-) [imported] - Streptococcus pneumoniae (strain R6)

C;Species: streptococcus pneumoniae

C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001

C;Accession: G98126

R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; Ede, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M; Y, P.; Sun, P.M.; Winkler, M.E.; Bacteriol. 183, 5709-5717, 2001

A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; A;Accession: G98126

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-441 &lt;KUR&gt;

A;Cross-references: GB:AE007317; PIDN:AAL00844.1; PID:915459750; GSPDB:GN00174

C;Genetics:

A;Gene: comp

C;Keywords: phosphotransferase

RESULT 6

G98126

histidine kinase (AgrC from Staphylococcus) homolog lmo0050 [imported] - Listeria

C;Species: Listeria monocytogenes

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C;Accession: AC1081

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blœc ; Domínguez-Bernal, G.; Duchaud, E.; Durand, L.; Bussurget, O.; Entian, K.D.; Fsihi, D.; Jones, L.M.; Karst, U. Science 294, 849-852, 2001

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitouram, A.; Ok, C.; Schlueter, T.; Simoes, N.; Tierre, A.; Vasquez-Boland, J.A.; Voss, H.; Wehla

A;Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:2153729; PMID:11679669

A;Accession: AC1081

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-431 &lt;GLA&gt;

A;Cross-references: GB:NC\_003210; PIDN: CAC98265.1; PID:916409409; GSPDB:GN00177

A;Experimental source: Strain EGD-e

C;Genetics:

A;Gene: lmo0050

Query Match 19.3%; Score 418.5; DB 2; Length 431;

Best Local Similarity 26.8%; Pred. No. 2.6e-18; Mismatches 142; Indels 111; Gaps 15;

Matches 125; Conservative 88; MisMatches 116; Conservatve 84; Indels 77; Gaps 10;

Qy 21 LFLFLSKVSYTLRKET--LPSISNPLIMAVTMVNWFYPAE--PFLFIAL---S 72  
 Db 4 IMATIOTQIFIAQILRNKVISKEGLVTTIAMLAFLPFTLVQWWSMIFVLVFSA 63  
 Qy 73 IYLNRON---SISLNITYGLPVASSDLEFRRAIFLFDGQIYWGSSITTYMEF 127  
 Db 64 LYWNKNVNVVASTINWVII-LUTIDS----IVGPFL-----VPG----- 99  
 Qy 128 AGIALSYLPSVENDIGRLKDSLTKMKVKRIP-----MNTIMLYLQVIVI 180  
 Db 100 ---ENPKYDEIN-----ELPPFLYCAGMANLVLVFSFLRKL--I 136  
 Qy 181 ESYVNPVTFKRFKVIVLILFLISI-----LSQYI-----LSQYI----- 214  
 Db 137 ERVNLSRFVRHKVYIISFIVSVALTVLFYMFNYIAGSIAGFDGSVLKINTLIFTGTYILL 196  
 Qy 215 -----KVKVQEMMAKEAQTRNIQYQSOIESLYKDRISFRDNLNTSL 261



R; Kuroda, M.; Ohta, T.; Uchiyama, I.; Bala, T.; Yuzawa, H.; Kobayashi, T.; Cui, L.; Quoc  
ma, A.; Mizutani-Otto, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001.

A; Title: Whole genome sequencing of metacillin-resistant Staphylococcus aureus.  
A; Reference number: A89758; MUID:21311952; PMID:11418146  
A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-371 <XRE>  
A; Cross-references: GB:BA000018; IPID:913701832; PIDN:BA843125.1; GSPDB:GN00149  
A; Experimental source: strain N315  
C; Genetics:  
C; Gene: agrC

Query Match 16.7%; Score 362; DB 2; Length 428;  
Best Local Similarity 24.8%; Pred. No. 7e-15; C  
Matches 111; Conservative 84; Mismatches 178; Indels 74; Gaps 12;  
Best Local Similarity 23.5%; Pred. No. 1.5e-15; C  
Matches 101; Conservative 92; Mismatches 162; Indels 74; Gaps 8;  
Best Local Similarity 23.5%; Pred. No. 1.5e-15; C  
Matches 101; Conservative 92; Mismatches 162; Indels 74; Gaps 8;

QY 12 LLTVLTLFLFLFLFSKVSNVPLSKELTFSISNFLIMIAVTMWNVNLYFPA 63  
Db 39 IAVFPFIPITVIALYSDISNVYIPLSGYFYIKEKENDYLNDLRLCTTIFGCNY--- 94  
QY 64 EPLVFTALSIYLNRONSLSUNIFGGLPAssDfRRRAILFILDFGTQGJMGSSITR 123  
Db 95 -----LASNIMQOLP-----SNRTVGPM-----GIFLOI 119  
Db 120 FIEALVSIITIFFYKNNH-----NTKEVAKTSVSYLLIVILLISYAA 168  
QY 124 MTEFAGIALSYLPLSVNFDIGRKDSLTKMKKKRLLPMPNITMLYLIQLV--YVI 180  
Db 169 HYDA---YDHFLGIMFELIQTIVFVFL-FLMKTCRKYKQIEO-ELNLK 221  
QY 236 YSQQIESLVLKDIORSFRHDYVNLISLURIGENKDLASKEKTHQIQLERGHOLDQTRN 295  
Db 222 YTESLFOQQOISKFRHDYKNLLSFKEINTNNKTAALKQELEQYSNRYLDKGFDY 281  
Db 296 GHILANTONDAYRGIIISAKLEAQNKIAVNEVSSKIQIPEMILLPITITSLCNAIE 355  
Db 282 KALYNTHNEFVKSLIATKHQAKELNIECYCECOPKLDPIDVPIPDICRIRLGILDNAIE 341  
QY 356 RAFFESINPELOLAFKKNSSIVTENSONSKKEOIDSFKFKEVSYKGSNGIGLAKVN 415  
Db 342 AASECIEKILVLYQODDQIEFSKNTKSNSIGTQRKNISTKRGHSGLGLNTIE 401  
QY 416 LLEHYVKTSLOTSNNHHFLR-OLIIX 441  
Db 402 FNQKFNPVFTYQKQEESEFSVOLITK 428

RESULT 10

G56273 protein-histidine kinase homolog Sapk - Lactobacillus sake (strain Lb706) plasmid pLSA60  
C; Species: Lactobacillus sake

C; Date: 03-Oct-1995 #sequence\_revision 03-Oct-1995 #text\_change 15-oct-1999  
C; Accession: G56273  
R; Axelsson, L.; Holck, A.  
A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-428 <XRE>  
A; Cross-references: EMBL:248542; NID:9695615; PIDN:CAB88426.1; PID:9695616  
A; Experimental source: strain Lb674  
C; Genetics:  
C; Gene: sppK

Query Match 16.0%; Score 348; DB 2; Length 448;  
Best Local Similarity 26.3%; Pred. No. 5.3e-14; C  
Matches 121; Conservative 92; Mismatches 199; Indels 48; Gaps 15;

QY 8 LSNGLITYLTVLFLFLFSKVSNVPLSKELTFSISNFLIMIAVTMWNVNLYFPA 67  
Db 6 VSVSLMONFVAIILFLFLYRVIORKRIFTKRIID----ILIAITFSILYFISDSL 59  
QY 68 FIAL-----SYLNQN-----SLSNFYGLPAssDfRRRAILFIFI-LDGQGI 113  
Db 60 VAVMLRQWHPHQKENKIKUTDANLILIVLIVLVAUGTISOFITISKSDFONI 119  
QY 114 WGSSSITTMIEFAGIALSYLPLSVNFDIGRKDSLTKMK-VKKRLLPMPNITMLYI 172  
Db 120 LNNASDITL---GIFEAFLFDGLFET---LKNKPTVQELHQINOLEELESLEK-QYF 171



R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berger, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capriano, V.; Carter, N.M.; Chodat, A.; Erlich, S.D.; Elmer, S.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, E.; Nature 390, 249-256 1997

A;Authors: Fouger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galler, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Hoisappel, S.; Hosono, S.; Hull, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Laufer, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel, Y.; M.; Ogawa, K.; Oduga, A.; Outeiga, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Scheifele, S.; Schroeter, R.; Scorfone, F.; Sekiguchi, J.; Sekowska, A.; Seron, Akeuchii, M.; Tamachi, A.; Tanaka, T.; terpstra, P.; Toogood, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wiat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zunstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. A;Accession: A69580; MID:98044033

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-542 <DNA>

A;Cross-references: GB:99108; GB:AL009126; NID:92633055; PID:CA12587.1; PID:92633082

A;Experimental source: strain 168

C;Genetics:

A;Gene: cikS

C;Superfamily: two-component sensor histidine kinase; sensor histidine kinase homology

Query Match 8.0%; Score 174; DB 1; Length 542;

Match	Best Local Similarity	Score	DB	Length
Matches	97;	Conservative	73;	Mismatches 134; Indels 98; Gaps 21;
Qy	35	KKEELFLSISNFLIMAVINVNVLVYPAPEPLFIALSIYLN--RQNSLSUNIFYGLPV	92	
Db	171	KKHLRLNLSSAVLVLW-----I-----LGFGAAVLAKSIRKDPL-----GLEPH	209	
Qy	93	ASSDIE--RRALEIFLDGCGTQIGVGSSTTTMIEFAGLALYLESPVENDIGRKDS	150	
Db	210	ETAAALYRERINMLFAIREG-----IAT---NRGV-----VIMMYNSAEM--	248	
Qy	151	LTKMKVKKRJIPMNNTMLLYLQLVLYTESVNVPIVKFRKEVIVVLLFLILISFL	150	
Db	249	--LKLPEPPTVHLRPTDVMG--AGIMSLTEKGMLNPQE----VSNDQFVLI----	293	
Qy	211	SQITKVKVQE-----IMAKQAQNITIQSQSOTESLYKDIRSFRRDYLN---IL	258	
Db	294	---NIKVMMHQQGGAYGIVSFREKTELKKLDRTEVRYKSSEDLRAGTHEPSNKVAIL	349	
Qy	259	TSLRIGIENNDLASTEKIHKIQLERKTGHOLDQDTRYNIGHANITQNDAYKGIIASAKLFAQ	318	
Db	350	GLELGEYDEDAIDLQKKEY--A-TONEHQD-----LPH--NIHSQVQAIALLGKISKAS	399	
Qy	319	NRIKIAVW-EVSSKTLQP-----FEMELDFTITLSTICDNATAAFAFSLNPELQIAFPFKNGSI	376	
Db	400	EKKVQVVIDNSSLAPLPAHIGLSPUUTIGNLIDNAEFAVAEQSVKEY-LFIFTDQHD	458	
Qy	377	VFIIONST-----KERQDVKSKTRENKSTKGSNRGIGLAKV	413	
Db	459	IVIEWSDTGVPPEK--LEAVFERGYSSKGMRGGLANV	497	

RESULT 15

E83699

two-component sensor histidine kinase BH0397 [imported] - *Bacillus halodurans* (strain C-C)

C;Species: *Bacillus halodurans*

C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C;Accession: E83699

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hirai, Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliophilic bacterium *Bacillus halodurans* and its relatives

A;Reference number: A83650; MID:20512582; PMID:11050132

A;Accession: E83699

A;Status: preliminary

A;Molecule type: DNA

A; Residues: 1-532 <STO>  
 A; Cross-references: GB:AP001508; GB:BA000004; NID:910172890; PIDN:BA04116.1; GSPDB:G  
 A; Experimental source: strain C-125  
 C; Genetics:  
 A; Gene: BH0397  
 C; Superfamily: two-component sensor histidine kinase; sensor histidine kinase homolog  
 Query Match 7.9%; Score 171.5; DB 2; Length 532;  
 Best Local Similarity 24.0%; Pred No 0.0035;  
 Matches 53; Conservative 44; Mismatches 87; Indels 37; Gaps 5;  
 Qy 218 VONEI~~MA~~QFAQIRITQ~~S~~QQTSL~~-----~~YKDISSFRHYOLNTLSLULGIENKDLA 271  
 db 299 IVNDEVGAGAISTFRULTEIKQQAKQLTGKVLYVEALRAQSHEFMNKLVHLGMVKTESYD 358  
 Qy 272 SIEKYHQI~~E~~KTGHQLODQTRYNIGHTAMINQDAVKGILSAKLEQPNKKLAVNVESSK 331  
 db 359 ELDNYI~~TH~~QINHRSTELNRV~~-----~~IKRIDS~~V~~LAFGLKGKSYAREKHTIDVQTKSV 412  
 Qy 332 IQLPENELL~~-DF~~ITLSTCUDNAIAAEFSLNPEIQLAFLAFKFKNGSIVFIQN~~ST~~KERQI 389  
 db 413 IPKASSEQMH~~ELT~~WILTGILNLIDNALEAVKA~~-----~~KEKTVLVOLEYSNRLHI 461  
 Qy 390 DV~~-----~~SKIFENYNSPKGSNRSIGLAKVNHILE 418  
 db 462 RVQDGPGIPDGEODIFKKGYSTKGENRGYGLYLVQSV 502

Wed Jul 31 08:30:58 2002

us-09-833-017-4.rpr

Copyright (c) 1993 - 2000 Compugen Ltd.	GenCore version 4.5
OM protein - protein search, using sw model	
Run on:	July 30, 2002, 15:20:52 ; Search time 15:86 Seconds (without alignments)
Scoring table:	US-09-833-017-4 BLOSUM62 Gapop 10.0 , Gapext 0.5
Searched:	105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters:	105224
Minimum DB seq length:	0
Maximum DB seq length:	200000000
Post-processing:	Minimum Match 0%, Maximum Match 100% Listing first 45 summaries
Database :	Swissprot_40;*
Pred.	No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.
<b>SUMMARIES</b>	
Result No.	Score Match Length DB ID Description
1	151 6.9 543 1 DCUS_ECOLI P39272 escherichia
2	139 6.4 428 1 KIMB_BACSU Q08430 bacillus su
3	129.5 6.0 532 1 DCUS_BACFD YE21_ARCFD 028851 archaeoglob
4	129 5.9 880 1 DPB_ECOLI P77510 escherichia
5	128.5 5.9 552 1 CHG_RHINE P72292 rhizobium
6	128 5.8 577 1 DCTS_BACSU P96501 bacillus su
7	126.5 5.8 535 1 ICN3_LACUA P37608 lactococcus
8	126.5 5.8 691 1 NTP2_MSPEV 09Y406 melanoplus
9	125.5 5.8 938 1 V120_HSV7J P52438 human herpe
10	120.5 5.5 430 1 SURA_BACAU P57240 buchnera ap
11	118 5.4 1024 1 RP0_PFLA P21421 plasmidum
12	117.5 5.4 214 1 SP2M_BACSU P37873 bacillus su
13	117.5 5.4 388 1 IF02_AQIRE 067472 aquifex aeo
14	116 5.3 656 1 YC25_POREU P51392 porphyra pu
15	116 5.3 850 1 DPO1_ANATH 059156 anaerococcus
16	115.5 5.3 451 1 YGYX_HAEIN P45316 haemophilus
17	115.5 5.3 873 1 COXI_ACACA Q37370 acanthamoeb
18	114.5 5.3 600 1 NIST_LACUA 003203 lactococcus
19	114.5 5.3 904 1 TLR3_HUMAN 015055 homo sapien
20	114.5 5.2 366 1 YP976_META 058366 methanococc
21	113 5.2 382 1 YC81_MESWT Q9mmp3 mesotrichia
22	112.5 5.2 883 1 YH16_YEAST P38781 saccharomyces
23	112.5 5.2 1 EX5B_BUGAI P57529 buchnera ap
24	112.5 5.2 496 1 IFHK_ECOLI P52101 escherichia
25	111.5 5.1 600 1 MUS2_BORLU 051125 borrelia bu
26	111.5 5.1 556 1 NU2M_PODAN P15578 podospora a
27	111.5 5.1 885 1 RA5C_ARCFD 029230 archeoglob
28	109.5 5.0 440 1 CAPE_STAU P38781 staphylococ
29	109.5 5.0 1274 1 BXF_CLOBO P30996 clostridium
30	109.5 5.0 1121 1 ALPI_SCHEO 010197 schizosacch
31	108.5 5.0 455 1 TRME_LACUA 09cd8 lactococcus
32	108.5 5.0 614 1 YDNK_LACIC P42377 lactococcus
<b>ALIGNMENTS</b>	
RESULT	1
ID	DCUS_ECOLI
NAME	STANDARD
PRT	543 AA.
AC	P39272; P76795;
DT	01-FEB-1995 (Rel. 31, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	Sensor protein dcus (EC 2.7.3.-).
GN	DCUS OR B4125 OR Z5727 OR ECSS5107.
OS	Escherichia coli, and
OS	Escherichia coli O157:H7.
OC	Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae; Escherichia.
OX	NCHI_TaxID=562, 83334;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=K12 / MO1655;
RX	PubMed=9533432; PubMed=7610040;
RA	Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L., Blattner F.R.; Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.8 through 100 minutes.; Nucleic Acids Res. 23:2105-2119(1995).
RA	SEQUENCE FROM N.A.
RA	STRAIN=O157:H7 / ED933 / ATCC 700927; PMID=21074935; PubMed=11206551;
RA	perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Evans P.S., Gregor J., Kirrpatick H.A., Postai G., Hackert J., Klink S., Boutin A., Shao Y., Miller L., Grotbeck E.J., Davis N.W., Lim A., Bimelaite E.T., Potamousis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.; Nature 409:529-533(2001).
RA	SEQUENCE FROM N.A.
RA	STRAIN=O157:H7 / RIMD 0509952; PMID=21156231; PubMed=11258796;
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Kuida T., Tamaki H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., Kubota S., Shioya T., Hattori M., Shinagawa H.; "Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12."; RT [4]
RA	CHARACTERIZATION.
RA	STRAIN=K12 / MO14100; PMID=9844044; PubMed=9765574;
RA	DNA 8.11-32(2001).
RA	CHARACTERIZATION.
RA	CHARACTERIZATION.
RA	"Functional regulation of gene expression in Escherichia coli by the two-component regulatory system"; J. Bacteriol. 180:5421-5425(1998). [5]
RA	CHARACTERIZATION, AND TOPOLOGY.



DR	InterPro; IPR003594; HATPase_c.	01-MAR-2002 (Rel. 41, Last annotation update)
DR	InterPro; IPR004359; HIS_KIN_sig.	DE Probable C4-dicarboxylate sensor kinase (EC 2.7.3.-).
DR	InterPro; IPR003661; His_kinA.	DE DCTS OR BH2752.
DR	pFam; PF02518; HATPase_c; 1.	OS Bacillus halodurans.
DR	pFam; PF00512; signal_1.	OC Bacteria; Firmicutes; Bacillus/Clostridium group;
SMART	HISPASe_c; 1.	OC Bacillus/Spaerulococcus group; Bacillus.
SM00388	HISPASe_c; 1.	OX NCBI_TaxID:86665;
PS00387	HIS_KIN; 1.	RN [1]
SS00387	HIS_KIN; 1.	RP SEQUENCE FROM N.A.
STRAINC-125	/ JCM 9153;	RC MEDINE=20512582; PubMed=11059132;
JCM 9153		RX Takiuchi H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
PubMed=11059132		RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA		RA Horikoshi K.
RN		RT "Complete genome sequence of the alkaliophilic bacterium Bacillus halodurans and genomic sequence comparison with <i>Bacillus subtilis</i> *;
[1]		RT Nucleic Acids Res; 28:4317-4331(2000)
NCBI_TaxID:86665;		RN -I- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM DCTS/DCTR.
[1]		CC PROBABLY ACTIVATES DCTR BY PHOSPHORYLATION. ESSENTIAL FOR EXPRESSION OF DCTP (BY SIMILARITY).
TRANSMEM	120	CC -I- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
TRANSMEM	121	CC -I- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
TRANSMEM	132	CC -I- SIMILARITY: CONTAINS 1 PAS (PER-ARMED) DIMERIZATION DOMAIN.
TRANSMEM	133	CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@ib-sib.ch).
TRANSMEM	153	CC ---
TRANSMEM	154	CC ---
TRANSMEM	166	CC ---
TRANSMEM	167	CC ---
TRANSMEM	187	CC ---
TRANSMEM	188	CC ---
TRANSMEM	218	CC ---
MOD_RES	221	CC ---
CONFFLICT	44	CC ---
SEQUENCE	328	CC ---
AA	428	CC ---
MW	47812	CC ---
MMW	52FBECAG6864F7040	CC ---
CRC64;		CC ---
Query Match	6.4%	Score 139; DB 1; Length 428;
Best Local Similarity	20.5%	Pred No. 0.16; : Pred. No. 0.16; : Indels 86; Gaps 17;
Matches	93	Conservative 86; Mismatches 188; : Indels 86; Gaps 17;
QY	7	ILSNGLTYITVVLFLPFKSVNVNTLSKKELTLEFSISNLIMIAVMVN-NLILYPAEP 65
Db	3	IIKDYLLHICFILPILKIQV--FWLGKPATIPVKNSGLVTLFAGASVCLIFPIHE 59
QY	66	LYFTALSTIYINRQNQLSNTFYGL--LFLASSDLERRAIFIPLDTQGIVMGSSITY 123
Db	60	MDYIQ----YGLQMPV-----TICFLYIYSTASGLVVAASVLCFE 95
QY	124	MIEFGIALSYLSEFSVNFIDGI--KDSLTKVKKRLLPNNITMLYLILQLYVIE 181
Db	96	LFLYEPSAMKEVTFVPLFLITIPILQPKWPNSKAKLLSLISCVETIFLFFASWILS 155
QY	182	SYNVIPTLKFKEFWIVLIL--LFLLISLQSQYKQKQVNEIMAQKEQKQRNQTOYS 237
Db	156	ALNI---LNFQKSGFIVYEAVGSLFRRSVLILSIVITESAENI----ALRSOLHS 206
QY	238	QOIESIKDTRSFRHDYLNLTSIR-----LGTEKNDIASIKIYHOILEKTGHOL 288
Db	207	EKMIVSELASAVHEVRNPILTVRGPVQOLLNEDTQKNSADYKLVLSLDRAQ-- 263
QY	289	QDTRNIGHANIONDAVGI----LSAKTLEAQNKI----AVRNEVSKRQLP 335
Db	264	-----GINTVNDMAQOLQYKEVEDLSALLKETSSLMYSANYKSVTVEAE--P 313
QY	336	EMELIDFTTISLICDNATEAAFESL--NPEQOLAFKKNSIVE--IIONSTKQID 390
Db	314	DLLIGDAKMKQKVINMKNSTEAVPHGMTHIS-AKRNGHTMINITINGVMTDHQ 372
QY	391	VSKIRENNTSKGSNRGICLAKYVNHILEYPKT 423
Db	373	MQLKGEPIVYSLKINGTGLGLTVFSIIEHHHGQT 405
RESULT	3	
DCTS_BACHD		
ID	DCTS_BACHD	STANDARD; PRT; 532 AA.
AC	09K97;	
DT	16-Oct-2001 (Rel. 40, Created)	
DT	16-Oct-2001 (Rel. 40, Last sequence update)	
Query Match	6.0%	Score 129.5; DB 1; Length 532;
Best Local Similarity	18.4%	Pred. No. 0.74; : Pred. No. 0.74; : Indels 159; Gaps 22;
Matches	97	Conservative 99; Mismatches 173; : Indels 159; Gaps 22;
QY	4	ALMILSNGLTITVLLFELLEFLSKVSNVTLKKEFT-----AEPYLFIASLYINRQNQLS 83
Db	24	ALMVSISLILGVYTSI--KDEDELSNRMTTAOLVAQHNTQDQWDAKPEEASRTLQPI 79
QY	40	--LFSSINFLIMIAVMVNVLNFY-----AEPYLFIASLYINRQNQLS 83
Db	80	VERIRVNDHYDITVLLGMDRIRITHIPTERLQTPFVGGDDEPAP--AEPYLFIASLYKATEGV 137
QY	84	NIFYGLFVASSDLERRAIFIPLDTQGIVMGSSITY--MTE----- 126
Db	138	VTVRAFMPLINGO--REOV-----GVAVGVSVLPSADMIQEFMQWPAALLGLITAL 186
QY	127	--FAG--TALSFLSFFNVNDIGRLKDSLK----- 153

		FT TRANSMEM	195	217	POTENTIAL.
	FT TRANSMEM	237	259	POTENTIAL.	
	FT TRANSMEM	268	290	POTENTIAL.	
	SQ SEQUENCE	880 AA;	102210 MW;	8AD9AB7B4E50671 CRC64;	
Db	187	FGFWGSWLLASHKRRQTFFNMEPEDELALHILVERDASNAIHEGVWAINKHEKITINNEAR	246		
Oy	154	-MKYKKRKLIPMMNTMLLYLTLVYVIESVYVPTLKFRFFFV--VVLFLF--	205		
Db	247	RMLGKKEAIGRN-----HEVIPDTKLPELISICKPLYQREFFIQR	289		
Oy	206	LISFSQSYTOKKVNNEIMA---OKEAQIRNTQYSQIESLYKDGRSFRRDYLWILTS	260		
Db	290	LV--FSNRPIQIDGETVGAIAFQDKSDVDRLEELTGQWFADALRVNHEYNSKLHT	347		
Oy	261	--IRLGTEENKDIASIEKIYHOTLEKGHQLDTRNIGHANIONDAVGILSAKILE	316		
Db	348	TAGLIQDDEGRKAQYIFDLEEQQEFSSVVMQ-----KHNDSLGLLGKVSR	397		
Oy	317	AQNKKAVNVEVSSK-IQLEP-MEILDFITISLCNAATEA--AFSLSNPOLAFFKK	372		
Db	398	GRELGVQVIEKDSFIDHPEGVTHDLVIVGNLIDNSDASSDQNKTVHVFGE	457		
Oy	373	NGSIVVILQNSTKEQIDV-SKIKENYSTKG-SNRGLAKVNHILE	418		
Db	458	NDFLKIRVRDNGEGIREVREKMFVRGRSTKSTSGRGIGFLQIAVE	505		
RESULT	4				
ID	Y221_ARCFU	STANDARD;	PRT;	880 AA.	
AC	028851;				
DT	01-MAR-2002	(Rel. 41, Last sequence update)			
DT	01-MAR-2002	(Rel. 41, Last annotation update)			
DE	Hypothetical protein AF1421.				
OS	Archaeoglobus fulgidus.				
OC	Euryarchaeota; Archaeoglobales; Archaeoglobaceae;				
OX	Archaeoglobus.				
RN	NCBI_TaxID=2234;				
[1]	SEQUENCE FROM N.A.				
RP	Y221_ARCFU	STANDARD;	PRT;	880 AA.	
RX	STRAIN=VC-16 / DSM 4304 / ATCC 49558;				
RA	028851; PMID=9389475;				
RA	Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Richardson D.L., Kerlavage A.R., Graham M., Hickey E.K., Peterson J.D., Ketchum K.A., Dodson R.J., Gwynn M., Karpas N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Recht C.I., McNeil L., Badger J.H., Glodek A., Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T., Cotton M.D., Spirogs T., Artlach P., Kaine B.P., Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;				
RT	"The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon <i>Archaeoglobus fulgidus</i> .";				
RT	Nature 390:364-370(1997).				
CC	-- SIMILARITY: BELONGS TO THE UPF0182 FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to licensee@isb-sib.ch).				
CC	EMBL; AE001005; AAB89825; 1; .				
DR	TIGR; AF1421; -. Hypothetical protein; Transmembrane; Complete proteome.				
FT	TRANSMEM 19 POTENTIAL.				
FT	TRANSMEM 61 POTENTIAL.				
FT	TRANSMEM 103 POTENTIAL.				
FT	TRANSMEM 152 POTENTIAL.				
RESULT	5				
ID	DPB_ECOLI	STANDARD;	PRT;	552 AA.	
AC	DPB_ECOLI				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Sensor kinase dpb (EC 2.7.3.-) (Sensor kinase cta).				
GN	DPB OR CITA OR MPDB OR B0619.				
OS	<i>Escherichia coli</i> .				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.				
OX	NCBI_TaxID=562;				
RN	[1] SEQUENCE FROM N.A.' AND CHARACTERIZATION.				
RC	STRAIN=K12 / MG1655; MEDLINE=9367124; PubMed=9741802;				
RA	Inger H., Miller C., Cohen S.N.;				
RT	"Destabilized inheritance of pSC101 and other <i>Escherichia coli</i> plasmids by DphA, a novel two-component system regulator.";				
RT	Mol. Microbiol. 29:49-59(1998).				
RN	[2]				



RP	SEQUENCE OF 1-112 FROM N.A.	446 KONPRASTIVSGYELRIGQITNLLENA----RSFVPR-----QNGRIVVRLLRSRL 493
RC	STRAIN=SU47 / 1021;	
RX	MEDLINE=9603283; PubMed=7559334;	QY 375 -SIVEFIQNSTKEQDVKSKIFKVNSTK-----GSNGIGLAKVNHLHYPKTSLOT 427
RA	Oesterraas M.; Stanley J.; Finan T.M.;	"Identification of Rhizobium-specific intergenic mosaic elements within an essential two-component regulatory system of Rhizobium species.";
RT	J. Bacteriol. 177:5485-5494(1995).	
RL		
CC	-!- FUNCTION: MEMBER OF A TWO-COMPONENT REGULATORY SYSTEM	
CC	CHVGLEXOS/YCHT INVOLVED IN REGULATING THE PRODUCTION OF SUCCINOCYAN ACTIVATES CHYT BY PHOSPHORYLATION.	
CC	-!- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.	
CC	-!- SUBUNIT: HOMODIMER.	
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.	
CC	-!- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@ib-sib.ch).	
CC	-----	
CC	EMBL: AF027298; ABH96631; -	
DR	AL59178; CAC4130; ALT_INIT.	
DR	U33941; AAB07686; -	
DR	InterPro: IPR04458; BCTR SENSOR.	
DR	InterPro: IPR03650; HAMP.	
DR	InterPro: IPR03554; HATPase_c.	
DR	InterPro: IPR04359; HIS_KIN_sig.	
DR	InterPro: IPR03651; HIS_KINA.	
DR	Pfam: PF00672; HAMP; 1.	
DR	Pfam: PF02518; HATPase_c; 1.	
DR	Pfam: PF00512; signal; 1.	
DR	SMART; SM0038; HATPase_c; 1.	
DR	SMART; SM0038; HSKA; 1.	
DR	PROSITE: PS50109; HIS_KIN; 1.	
KW	Sensory transduction; transferase; kinase; phosphorylation; transmembrane; inner membrane; Exopolysaccharide synthesis; complete proteome.	
FT	DOMAIN 1 29 CITOPLASMIC (POTENTIAL).	
FT	TRANSMEM 30 50 PEPTIPLASMIC (POTENTIAL).	
FT	DOMAIN 51 260 PEPTIPLASMIC (POTENTIAL).	
FT	TRANSMEM 261 281 PEPTIPLASMIC (POTENTIAL).	
FT	DOMAIN 282 577 CITOPLASMIC (POTENTIAL).	
FT	DOMAIN 347 575 HISTIDINE KINASE (BY SIMILARITY).	
FT	MOD_RES 350 350 PHOSPHORYLATION (BY SIMILARITY).	
FT	CONFLICT 136 136 L -> P (IN REF. 1).	
FT	CONFFLICT 266 266 V -> A (IN REF. 1).	
FT	SEQUENCE 577 AA; 63720 MW; BEF2DD4AFOE2B15 CRC64;	
Query Match	5.8%; Score 127; DB 1; Length 577;	
Best Local Similarity	22.5%; Pred. No. 1;1;	
Matches	68; Conservative 57; Mismatches 89; Indels 88; Gaps 14;	
Qy	197 IVYIILFLILISPLRSQYKQ-----KVONEI--MAQEQAQINITORSQQIE 241	
Db	272 LVNVILSLSLSSIANPRLSAAIRVRRGAKEREIPPFSSRDEIGNLSVALREMT 331	
Qy	242 SLYKD---IRSF---RHGYMLTSLRGLENKDLASIKIYHOLETGHHQDTRY 293	
Db	332 TALYDRIAAIENFAADVSHELKNPLTSRSAVETLPLARNEESKKRIMDVLIQHDRV--- 387	
Qy	294 NIGHLANIONDAYKGILSAKILEAQNKTKAVN-----VEVSSKIQ-LPMEMLDFIT-- 344	
Db	388 RDLRLISDISDAS--LDAELARAKKVDLKLGLDVLVEISRQTRGSKKPVVLDFVDR 445	
Qy	345 -----ILSLCDWAIAAREFSLNPEIQLAFFKNG----- 374	
RESULT	7	
DCTS_BACSU		
ID	P96601	STANDARD; PRT; 535 AA.
AC		
DT	16-OCT-2001 (Rel. 40; created)	
DT	16-OCT-2001 (Rel. 40; Last sequence update)	
DE	01-MAR-2002 (Rel. 41; Last annotation update)	
GN	Probable C4-dicarboxylate sensor kinase (EC 2.7.3.-).	
OS	Bacillus subtilis.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacilli; Bacillus/Staphylococcus group; Bacillus.	
OX	NCBI_TaxID=1423;	
DR	[1]	
RN	SEQUENCE FROM N.A.	
RC	STRAIN=168;	
RX	MEDLINE=98000887; PubMed=9341680;	
RA	Beilin C., Ayora S., Exley R., Hirschbein L., Ogasawara N., Kasahara Y., Alonso J.C., Le Regard F.; "Characterization of an lrp-like (lrcp) gene from <i>Bacillus subtilis</i> ."; Mol. Gen. Genet. 256:63-71(1997).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=168;	
RA	Kurst F., Ogasawara N., Yoshikawa H., Danchin A.; "Regulation of the transport system for C4-dicarboxylic acids in <i>Bacillus subtilis</i> ."; Microbiology 146:263-271(2000).	
RT	-!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM DCTS/DCTR. PROBABLY ACTIVATES DCTR BY PHOSPHORYLATION (BY SIMILARITY). ESSENTIAL FOR EXPRESSION OF DCTR.	
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).	
CC	-!- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.	
CC	-!- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@ib-sib.ch).	
CC	-----	
DR	EMBL: AB00148; BAA1928; -	
DR	EMBL: 299106; CAR12252; -	
DR	SubList; BG12073; dctsS.	
DR	InterPro: IPR003594; HATPase_c.	
DR	InterPro: IPR004359; HIS_KIN_sig.	
DR	InterPro: IPR000014; PAS.	
DR	Pfam: PF02518; HATPase_c; 1.	
DR	Pfam: PF00512; signal; 1.	
DR	SMART; SM0038; HATPase_c; 1.	
DR	PROSITE: PS50109; HIS_KIN; 1.	
DR	PROSITE: PS50112; PAS; FALSE_NEG.	
KW	Sensory transduction; Transferase; Kinase; Transmembrane;	

KW Phosphorylation; Complete proteome.  
 FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 12 32 POTENTIAL.  
 FT DOMAIN 33 172 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 173 193  
 DOMAIN 194 535  
 DOMAIN 213 276 HISTIDINE KINASE.  
 FT DOMAIN 333 528 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 MOD-RES 336 336 SEQUENCE 535 AA; 59942 MW; 61BDABBE8F1OCFE CRC64;

## Query Match

Score 5.8%; DB 1; Length 535;

Best Local Similarity 18.7%; Pred. No. 1.1; Mismatches 137; Indels 137; Gaps 24;

Matches 84; Conservative 91; Mismatches 137; Indels 137; Gaps 24;

QY 63 ADEPVIFALSIYLVRNRLSUNIYGLPVPASSDILFRRAIFFTLDGTQGIVNGSSITT 122  
 Db 121 ARIHYPE---AKGEITGAVRFAV---PVKDDQDLNQGV----VLEGKTU--- 160  
 QY 123 YMTEFAGTA--ISYL---FLSVENFDIGRLKDSLTKMKVKRLIPNMITMLYYLIQ 175  
 Db 161 ----PGIADILHLHKRDIAPITVLTGIGLAGFELLARHIKO----- 199QY 176 VVYVIESYNV----PTLKPKRFVIVVLLTFLTISLSQTSQKQ--KVQNEIMAQ--- 225  
 Db 200 -MEQLERHEITVERMEERTATPFSMNEGVIADNLKVITIFNERAKQFEVGQELIGKV 258

QY 226 --KEOI----RNTOYSQIE-----SLYKD----- 246

Db 259 EVLKDSDLPEIWERKAVYNEEIRVSGKVIMSRIPVKKVIGAVAIFQDTEAKMA 318

QY 247 ----TNSF----RHDVNTIUTS---LRQIENKLASIEKIVHOILETGHQDQ 290  
 Db 319 EELTGIVNFEALRVONHEMKLFTIAQLIQLGSEKAL-----QLAFOAST 366QY 291 TRYNGIHL--ANTONDVAKGTLISAKLEONKKAIVNVVSSKQ-QDQE-MEDLFETIL 346  
 Db 367 EDENVTTEFLHRKSIONDAAGILSKIRRGRELGIHAVIDENSSIQQFPEHVQDHDTWLL 426

QY 347 STLCDNATEAFESLSPRE--IOLAFFKKGSYFVIIQ-NSTREKQDVKPFKENTYS-T 401

Db 427 GLNIENAF-GSEPIVQEDRKADISQTDDILAILIEDNGCGJIEPTHMPLYDKGFTVN 485

QY 402 KSNNRGIGLAKNHIEHYHPKTSQNSNH 430

Db 486 KTTGGTYGYLYLVQQLIDKGSGT-IEVDSH 513

RESULT 8

LCN3 LACTA STANDARD; PRT; 691 AA.

AC P37608; DT 01-OCT-1994 (Rel. 30, Created)  
 01-OCT-1994 (Rel. 30, Last sequence update)  
 01-MAR-2002 (Rel. 41, Last annotation update)

DE Lactocin 481/lactococin transport ATP-binding protein lcnr3.

GN LCNR3.

OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis)

OC Firmicutes; Bacilli/Clostridium group; Streptococcaceae;

CC Lactococcus  
 NCBI\_TAXID=1360;  
 [1]

RP SEQUENCE FROM N\_A.

RC STRAIN-ADRIA 851030;

MEDLINE-9428641; PubMed-8017645;

RA Rince A., Dufour A., le Pogam S., Thuault D., Bourgeois C.M.,

Pennec J.P.; Cloning, expression, and nucleotide sequence of genes involved in production of lactococcin DR, a bacteriocin from Lactococcus lactis

RT subsp. lactis; Appl. Environ. Microbiol. 60:1652-1657(1994).

RL CC -1- FUNCTION: PROBABLY IMPLICATED IN THE EXPORT PROCESS OF THE

KW Phosphorylation; Complete proteome.

FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 12 32 POTENTIAL.  
 FT DOMAIN 33 172 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 173 193  
 DOMAIN 194 535  
 DOMAIN 213 276 HISTIDINE KINASE.

MOD-RES 336 336 SEQUENCE 535 AA; 59942 MW; 61BDABBE8F1OCFE CRC64;

CC CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. HYLY SUBFAMILY.CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation and the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).CC CC DR EMBL: 941581; AAC72259\_1; -.  
 DR DR InterPro: IPRO0359; AAA.  
 DR DR InterPro: IPRO0110; ABC\_transporter\_tmem.  
 DR DR InterPro: IPRO0349; ABC\_transporter.  
 DR DR InterPro: IPRO1687; ATP\_GTP\_A.  
 DR DR Pfam: PF000664; ABC\_membrane; 1.  
 DR DR SMART: SM00382; AAA; 1.  
 DR DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 DR DR Transport; Protein transport; Bacteriocin transport; Transmembrane; ATP-binding.  
 KW KW FT TRANSMEM 157 179 POTENTIAL.  
 FT TRANSMEM 189 208 POTENTIAL.  
 FT TRANSMEM 262 284 POTENTIAL.  
 FT TRANSMEM 289 311 POTENTIAL.  
 FT TRANSMEM 381 403 POTENTIAL.  
 FT FT NP\_BIND 497 504 ATP (POTENTIAL).  
 SQ SEQUENCE 691 AA; 79934 MR: 427600778ECOB47 CRC64;

Query Match Score 5.8%; DB 1; Length 691;  
 Best Local Similarity 19.5%; Pred. No. 1.5; Mismatches 179; Indels 149; Gaps 25;  
 Matches 103; Conservative 96; Mismatches 179; Indels 149; Gaps 25;

QY 14 TYLTVLFLF-----FSKVSNUVNTSKKELTLSISNFNLIMAYMV----- 55  
 Db 158 TFL-VKFSLFLISQVALWFSTTLDINKSHDITY--FIMMISVLVOTLSLMLGA 213

QY 56 --NNVLFVYAAPEVIFALSIYLVRNRLSUNIYGLPVPASSDILFRRAIFFPDLGTOI 113  
 Db 214 QRNINLY-----ESK1SRQPKFG-----IFSRPLLYF-RNSVGT 248

QY 114 VMGSSIIITWYMFAGIALSYLELSVFNVDIGRUDSKLSDTRKMKVKELIPNMITMLYYL 173  
 Db 249 IIEKINRIGIRD-GILIK-IFPSLNF-----FTVIVIYTGTIFTLFLV 297

QY 174 IQWVIVLWVNVITLKRKFWVIVLFLFILISFQYKQVNEIMAQEQVIRN 233  
 Db 298 MNLLMIFPSL-----SKRQANIOYQTQTFTSVVOEDLNQIOBKAGNEKECKVRW 354

QY 234 TQISOQIESIYKDIRSF-----RHDLNLTSLRGLT-ENKDLASIEKIVHOLE 282  
 Db 355 TKSQSIYIISYKINTLNGITSAFNQGNYCIVLMMIFGTYLQNLVZIPDL--IIF 411

QY 283 KTG-----HOLQDTRNIGHA-----NLONDAYVKGILSAKI-- 314  
 Db 412 QSGISLFLSYAVNQIOPDMTEISRSIYGNKISDLTENPORIDTEKHSSNAITLKD SY 471

QY 315 -LEAQN--KRIANVEWSSKIQIPEMELDFITIISLC-----DNAEAFESLNPEI 365  
 Db 472 SYELNYIYFNINNFSIYKGEKKAIVGKSGKSTNLNIGLISYEGVYGENLROI 531

QY 366 QIAFFKKN--GSTVFIQIINSTKEKQIDV-----SKIFKENYST 402  
 Db 532 GYVSDQMNFLRKGSLTENIYVNNNSEELDIOKINDVADNMELYDLSPOKIFSQLFEN 590

QY 403 GSNRGIG-----LAK--VNH--ILEHYPKTSQTSNHHILFKOLL 438  
 Db 591 GKNLSSGQIQORLIIAKSLINNNKEFWDPEPFSLDQNRRHIVKNL 637

RESULT	9	QY	184	- - - - - N V I P T L K E R F - - - W V I V Y - - - - -
ID	NTP2_MSPEV	STANDARD;	PRT;	717 AA.
AC	O9YW0R;			
DT	16-OCT-2001 (Rel. 40, last sequence update)			
DT	16-OCT-2001 (Rel. 40, last annotation update)			
DE	Nucleoside triphosphatase II (EC 3.6.1.15) (NTPase II) (Nucleoside triphosphate phosphohydrolase II) (NPH II).			
GN	MSV08B.			
OS	Melanoplus sanguinipes entomopoxvirus (MSPEV).			
OC	Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae; Entomopoxvirus B.			
OX	Entomopoxvirus B.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=TUSCON;			
RK	Medline=99102612; PubMed=9847359;			
RA	Afonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.I.; "The genome of <i>Melanoplus sanguinipes</i> entomopoxvirus.";			
RT	J. Virol. 73:533-552(1999).			
RL	73:533-552(1999).			
CC	DURING TRANSCRIPTION OF EARLY MRNAs, PRESUMABLY BY PREVENTING R-LOOP FORMATION BEHIND THE ELONGATING RNA POLYMERASE. ACTS AS NTP-DEPENDENT HELICASE THAT CATALYZES UNIDIRECTIONAL UNWINDING OF 3'-TAILED DUPLEX RNAs. MIGHT ALSO PLAY A ROLE IN THE EXPORT OF NEWLY SYNTHESIZED mRNA CHAINS OUT OF THE CORE INTO THE CYTOPLASM.			
CC	REQUIRED FOR PROPAGATION OF VIRAL PARTICLES (BY SIMILARITY).			
CC	-1- CATALYTIC ACTIVITY: NTP + H2O = NDP + phosphate.			
CC	-1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAH SUBFAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).			
CC	EMBL; AF063866; AAC07810.1; -.			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR002464; DEAH_AMP_helicase.			
DR	InterPro; IPR001650; Helicase_C.			
DR	pfam; PF00271; helicase_C; 1.			
DR	SMART; SMD090; HELIC_C; 1.			
DR	PROSITE; PS00690; DEAH_AMP_HELICASE; FALSE_NEG.			
KW	ATP-binding; Helicase; Hydrolase; Transcription.			
FT	DOMAIN			
FT	NP_BIND			
FT	SITE			
FT	SEQUENCE			
SO	717 AA; 84725 MW; E8FB04661F1EA863 CRC64;			
Query Match	5.8%; Score 125.5; DB 1; length 717;			
Best Local Similarity	18.3%; Pred. No. 1.8;			
Matches	107; Conservative 97; Mismatches 172; Indels 209; Gaps 24;			
QY	15 YITIVLFLFLFSKVNTVSKKEETLEFSISNFILMIAVTMVNLFPAAEPLFYALIY 74			
DB	26 YNYMAYVLFL---PNNTATNSYTKKEFEYVMOFAIAl---YFVKIWHINIC 75			
QY	75 LNR-----QNSLSNIFYGLPVASSDILRAIF---FILEG----- 109			
DB	76 LNRFYIYLNSBEFKNNISINTVNL-----LNNELKFEDDNILLINGKNLKISAYSY 128			
QY	110 ---TQGIVMSSSIITYMI-----EFGIALSYLFSVFNDVIGRKDSITKMKVKK 159			
DB	129 VTIISOTINISLNKQYQIGIESANYLGILPSYKONFYDKNL---FSPTKSELST 184			
QY	160 LIPMNITMLYL-----LIVQY---VYESY----- 183			
DB	185 MIDVQKIFEIFISKRNCLISGGIGKTVIPKLFWWNLFDGYEEWNUTSNENKNIND 244			
Query Match	5.8%; Score 125.5; DB 1; Length 938;			
Best Local Similarity	20.3%; Pred. No. 2.5;			
Matches	92; Conservative 70; Mismatches 112; Indels 179; Gaps 22;			
QY	2 NEALML-----SNGLLTYLFLFLFSKVNVNLFPAAEPLFYALIY 43			
DB	577 NRLLLILIFACKLMLPSYLNLL-----YLLMLHFT---LQFKVGLGHFSIHAITOKF 630			
QY	44 -----SNEJLMIAVTMVNLFPAAEPLFYALIY 87			
DB	631 DNINSLTQTFIPTKTFNLFLSLL-----AYTMHOTVNPWIKQTSEN1-- 676			
QY	88 GLPVAASSDLFRRRAIFFLGDGTQGIVMGSSITTYMIEFGIALSYLFE-----IS 138			

Db 677 -----ALLEKEYIDFTK--KCSSLATTCYLNLENFAVNHYFGKNKGSPSLS 721  
 QY 139 VNVNDIGRL-----KDSUTKKVKKRJTPMNTMLLYLTLQVLYIESYNVPTL-K 190  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 722 AYRTOTSKLIEFSKLEKDRDQEIKVSKTF--IEMI-----ONWVNENITK 764  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 QY 191 FRKFVIVVYLFLILSFLSQYTKQVKVNEIMAKEAQIRNTQYSSQESILYKDRSF 250  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 765 FFDLV-----SNQTLQNFII-----VERISSHANTTYQDV- 795  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 QY 251 RDYLNNTSLRGLENKLASITYHOILEKTHGHOLODTRYNIGHANIONDAYGI- 309  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 796 -----LNSIDCHF---SNMQLQSFKNIVVWIDVNL-TKNI 829  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 QY 310 --LSAKILEAQNKKLVNVYEVSSVTOPEMELDFITLISICDNAEAFFESLAPPEIQ 366  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 830 NESLASOLIEN--KKIVVKQDQTYNOLNVD---DFVTWIKSLUNNFEKQFTINERR 883  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 QY 367 LAFFKKNGSTVITIONSTKEOIDVSKIFEN- 399  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 884 -----FMLEGIPDQIQPFLDVFDERY 905  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 11

SURABUCAI	ID	RPOB_PLAFA	ID	RPOB_PLAFA
SURABUCAI	STANDARD;	PRT;	PRN;	PRN;
		430 AA.		1024 AA.

RC STRAIN=TOKYO 1988;  
 RX MEDLINE=20445177; PubMed=10993077;  
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
 RT "genome sequence of the endocellular bacterial symbiont of aphids  
 RT Buchnera sp. APS";  
 CC Nature 407:81-86(2000).  
 -!- SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AP001118; BAB13858\_1; -;  
 DR Interpro; IPR00297; Rotamase;  
 PFam; PF00539; Rotamase\_1.  
 DR PROSITE; PS01096; PPIC\_PPASE\_1; FALSE\_NEG.  
 DR PROSITE; PS01098; PPIC\_PPASE\_2; 2.  
 KW Isomerase; Repeat; Signal; Complete proteome.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 430 SURVIVAL PROTEIN SURA HOMOLOG.  
 FT DOMAIN 179 277 PPIC 1.  
 FT DOMAIN 286 385 PPIC 2.  
 SEQUENCE 430 AA; 50737 MW; CAA4C425B3FBA5788 CRC64;

Query Match 5 %; Score 120.5; DB 1; Length 430;  
 best Local Similarity 23.0%; Fred. No. 2;  
 Matches 75; Conservative 55; Mismatches 111; Indels 85; Gaps 19;

QY 168 LUYLLIOVLYIESINVPIKFRKEVWV-----WILFL-----ILISFL 210  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Db 7 IFFVSSIFPVVLAKNNQVDI----TATVDEILNSDWNEILVFLKKKKFIPLK 61  
 QY 211 SOYTKQKVNEIMAQ---KEAQIRNTQYSSQESILYKDRSFH--DYNLITLSRL 263  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 62 SDEKEKVLKLEKLDLILQEANSKNNITREQIDPVKIALKKHSUDHEKKQILR- 120  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 QY 264 GIENKDLAS--IJKIYHOILEKTHGHOLODTRYNIGHANIONDAYGI 321  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 121 NIKPSYFDNFIFKIEILMKT--IQL-YELHKRINISEQEVNTFK-KLIKDNKF 174  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 QY 322 IAVNEVESSKTOLPEMELDFTILSTICDNAIEAA-----FESTUNEI 365  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 175 KKINL--SYILLPSLKDSD-----DNVNRNTKIRENIVKLKGYDFPKLLEC 222  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 QY 366 QLAFFKKNGSTVITIONSTKEOIDVSKIFEN-NYSTK-----SRIGIGIAKUWH 416  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 223 E-----KINKS-TTIVKMMFWKPILDIONSFRTLNFFKKQOGLGPVKGKLYLKVDI 276  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 QY 417 LEHYPKTSLOTSH--HHLFKQLLI 440  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 Db 277 --HHRKENIVTEFYMQHCLIKPSVIL 300

RESULT 12

RPOB_PLAFA	ID	RPOB_PLAFA	ID	RPOB_PLAFA
RPOB_PLAFA	STANDARD;	PRN;	PRN;	PRN;
		1024 AA.		1024 AA.

RP SEQUENCE FROM N.A.  
 RC STRAIN=EW(C10);  
 RX MEDLINE=95107345; PubMed=7808472;  
 RA Gardner M.J., Goldman N., Barnett P., Moore P.W., Rangachari K.,  
 RA Strath M., Whyte A., Williamson D.H., Wilson R.J.M.;  
 RT "Phylogenetic analysis of the rpoB gene from the Plastid-like DNA of  
 RT Plasmodium falciiparum.,"  
 RL Mol. Biochem. Parasitol. 66:221-231(1994).  
 RN [2]  
 RP SEQUENCE OF 328-1024 FROM N.A.  
 RX MEDLINE=91180755; PubMed=2011147;  
 RA Gardner M.J., Williamson D.H., Wilson R.J.M.;  
 RA Gardner M.J., Williamson D.H., Wilson R.J.M.;  
 RT "A circular DNA in malaria parasites encodes an RNA polymerase like  
 RT that of prokaryotes and chloroplasts.";  
 RL Mol. Biochem. Parasitol. 44:115-124(1991).  
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC RNA(N)  
 CC -!- MISCELLANEOUS: THIS RNA POLYMERASE IS ENCODED ON A CIRCULAR DNA.  
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X75544; CAA53232\_1;  
 DR EMBL; X52177; CAA36427\_1;  
 DR EMBL; X95275; CAA61572\_1;  
 DR PIR; S10438; RNZOBF;  
 DR InterPro; IPR00512; RNA\_pol\_B;  
 DR Pfam; PF00562; RNA\_pol\_B; 1.



FT	TRANSMEM	129	151	POTENTIAL.
FT	TRANSMEM	158	175	POTENTIAL.
FT	TRANSMEM	179	196	POTENTIAL.
FT	TRANSMEM	203	225	POTENTIAL.
FT	TRANSMEM	304	326	POTENTIAL.
FT	TRANSMEM	347	369	POTENTIAL.
SQ	SEQUENCE	388 AA:	45529 MW;	F06E78EEE1B9941 CRC64;
Query Match		5.4%;	Score 117.5;	DB 1; Length 388;
Best Local Similarity		22.4%;	Pred. No. 2.6;	Mismatches 80; Conservative 1;
Matches				Indels 95; Gaps 15;
QY	2 NEALM----ILSNGLITYLVLFULFLRSKVSAWTLSKELT--LFSISNFLMI	94	94	
Db	7 NEVLYLVAVLVIASFISLISLFGEVILLLIVLYKILRKERKGSLUPGILLYSL-----	60	60	
QY	51 AVTMVNVLFYP-----AEPYFIFTALSIYLNRON-----SLSUNIFYGLPVAS	94	94	
Db	61 -TVLSTLAIFPKRFLKGIEBGLFOFYFLNUKKEVKGSKIFPKLUGSLILLPVF	118	118	
QY	95 SDLFRATTAIFFILDGT-QGIVMGSSITTMWI----EFAIASYLFLSVFNVDIGRL	147	147	
Db	119 YKFVYKGPPKPIWGTEVGFFYALLSITTEFLFFKERFVYIPIFLEFLAV-----	170	170	
QY	148 KDSLJKMKVKKRJLIPMNITMLLXLYLTQVLYVIESNVNIPFLKFRFVVVY-LILFLIL	205	205	
Db	171 -----IFLSARKSMMIAFFVFLYFLFLV-----KSKKLKGKLAWSVNFIL	213	213	
RESULT	15			
YC26 PORPU	STANDARD:		PRT;	656 AA.
ID	YC26 PORPU			
AC	P51392;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DE	Hypothetical sensor-like histidine kinase ycf26 (EC 2.7.3.-).			
GN	YCF26			
OS	Porphyra purpurea.			
OG	Chloroplast.			
OC	Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.			
OX	NCBI-TAXID=2787;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AVONPORT;			
RA	Reith M.E., Munholland J.;			
RT	"Complete nucleotide sequence of the Porphyra purpurea chloroplast genome."			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).			
CC	-1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.			
CC	-1- PAS (PER ARNT-SIM) DIMERIZATION DOMAIN.			
DR	EMBL; IPR004338; BCFRLSENSOR.			
DR	InterPro; IPR003660; HAMP.			
DR	InterPro; IPR003594; HATPase_c.			
DR	InterPro; IPR003539; HIS_KIN_Sig.			
DR	InterPro; IPR003661; His_kinA.			
DR	InterPro; IPR00014; PAS.			
DR	InterPro; PF00672; HAMP.			
DR	InterPro; PF02518; HATpase_c.			
DR	Pfam; PF00889; PAS.			
DR	Pfam; PF00512; signal_1.			
DR	PRINTS; PRO0344; BCFRLSENSOR.			
DR	SMART; SM01304; HAMP.			
DR	SMART; SM00387; HATPase_c.			
DR	SMART; SM00388; HISKA.			
DR	SMART; SM00191; PAS.			
DR	PROSITE; P50109; HIS_KIN.			
DR	PROSITE; P50112; PAS.			
FT	DOMAIN MOD_RES	424	659	HISTIDINE KINASE.
FT	SEQUENCE	427	427	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
Query Match		5.3%;	Score 116;	DB 1; Length 656;
Best Local Similarity		20.5%;	Pred. No. 5.9;	Mismatches 68; Conservative 68; Mismatches 68;
Matches				Indels 110; Gaps 15;
QY	161 IPDNNTMLLYLQLIQL-----WVIESNVIPLK--FRK-FWVIVYLIL--FLIL	206	206	
Db	340 LPEDIQQLPFLINDTRKNFQSCICEQICLQLQKVKFRFLVLTVIDHKVSLK	399	399	
QY	207 -TSFLSYVTKO-NEMAKEQAINTTQYSQIESYKDRSF---RHDYNL	258	258	
Db	400 GIAMTIDQRTDQEVELNEKQ-----FLSNVSHERTPLFNLNSFLTYEHSDIDS	453	453	
QY	259 TSLR-LGLENKDLASSTEKKYHOLERTGHOLQDTRYNIGHANTOND-----A	305	305	
Db	454 QKLEFLAIAKNTGRTRLVDL-----LSRLESDEQTYLOPTDVA	498	498	
QY	306 VKGILSKTAKLEONKKVIAVNEVSSKIQOLPEMELLDFTTFLSTLCDNAE-----	356	356	
Db	499 VEOTIRYQVLSAKDKRKLHDIEQNL--CVLGINNHLQILQILANLWVNSLKFTHPN	555	555	
QY	357 -----AFESLNPMEIOLAFFKRGNSVFTIONSTREKODVSKFK---ENYSTKGNS	405	405	
Db	556 IILRAYTVDDLKTEVQHENSQVKVREICDNGIGISRNQERFARFLRIENVHLEG	615	615	
QY	406 RGIGLAKNTHLEHYKTSLQTSNNHHFLKQ	437	437	
Db	616 RGGJLSVKNLQK-----NSETHLSEL	640	640	
Search completed: July 30, 2002, 15:20:54				
Job time: 365 sec				

Wed Jul 31 08:30:59 2002

us-09-833-017-4.rsp





```

DR EMBL; AJ005401; CAB54587.1; -.
DR EMBL; AJ274419; CAC18581.1; -.
DR EMBL; AJ275641.0; CAC03516.1; -.
DR EMBL; AE007364; AAKY4685.1; -.
DR TIGR; SP0527; -.
DR InterPro; IPR001064; Crystallin.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR Kinase; Complete proteome;
SEQUENCE; 446 AA; 1824 MW; D8BCD28F5DF274C CRC64;
SQ

```

DR	InterPro; IPR003594; HATPase_c.		
DR	InterPro; IPR004359; HIS_KIN_sig.		
DR	Pfam; PF00518; HATPase_C'-1		
KW	kinase; Transferase; Transmembrane.		
FT	42		
FT	TRANSMEM	62	POTENTIAL.
FT	78	98	POTENTIAL.
FT	TRANSMEM	142	POTENTIAL.
FT	179	199	POTENTIAL.
SEQUENCE	426 AA;	48387 MW;	CRC640 CRCFC9EC3C4C0

Query Match	Score	DB	Length
Best Local Similarity	25.1%	2	426;
30.0%	Pred. No.	5.7e-24	
Matches	131;	Conservative	Mismatches
Qy	18	VFLFELFESKSNVNLVKRTEFLSISNEMILATMVNVLPAAFLYFLASIVLN R	77
Db	11	VTSLLIYHTRISSIKLTMRELALVFLGTFESSILS--TEFAFWIMMLTAIGKD GL	66
Qy	78	ONSISLNIFYGLPVASSDIFERRATIFFLDGTOGIVMGS---SIITTYMEFAGIAL	132
Db	67	HOKKYIAFFYSVY-----SYTFSFGUNNLSLGTFVGENIYIPIFTALPVI	116
Qy	133	SYLFSLSVENWDIGRKLDSTKMKVKRLPNMNTMLYLQOLIVYTESYNVPTL--	189
Db	117	NELICKLINPPDFELNSN-GYATRNLVIGVN--FIFFLCTVQY-GSYWEDTGDI	170
Qy	190	-KFRKFVWVVLILFLILISPLSQTKQTONELMAQKAQIRNITOVSQESLYDIR	248
Db	171	GTVRYYLITGSFLILLIYLALVLYNLKLSAIOQSOLCRLREQIKRMEQTQSQENLYTEVR	230
Qy	249	SFRHDYUNILTSURGIGENDKLASETEKIQYOLETKGHQDQDPTYNGHLANTONDVAKG	308
Db	231	SFRHDYUNMWSLHSQGEDNEITEKTVWAEVLSGINASIQSSYKSLNLSKRVPAIKS	290
Qy	309	ILSAKILEAQNKLAIVNVEYSSKIQLOPEMELDFITLILCNOIAEAFESLNPEOLA	368
Db	291	TLESKTIYAQONGIEVNIETISDVUTYTFDLDYIRVIAFLDNOIAEASLSTDLPMEMSA	350
Qy	369	FRIKKNGGIVFIQIONSKERQIDVSKIFKEWYSTKOSNRIGIQLAKVNHLIEHYKTSQTS	428
Db	351	FTEEDDSQTIVLNNAAPDSYDKRRIFESGRSTKGNRGIGLATVKDILKNYPNASLETE	410
Qy	429	NHHFLQKOLLIK 441	
Db	411	FRNLSLTQKLAIK 423	

DR InterPro: IPR003980; Phosphopant\_attach.  
 DR Pfam: PF02518; HATpase\_c; 1.  
 DR SMART: SM00387; HATpase\_c; 1.  
 DR PROSITE: PS00012; PHOSPHOPANTETHINE; UNKNOWN\_1.  
 KW Kinase; Complete proteome; 448 AA; 52221 MW; 230C003A104CBEF5 CRC64;

Query Match 24.8%; Score 539.5; DB 16; Length 448;  
 Best Local Similarity 29.9%; Pred. No. 1.3e-23;  
 Matches 137; Conservative 105; Mismatches 173; Indels 43; Gaps 11;  
 Matches 137; Conservatve 105; Mismatches 173; Indels 43; Gaps 11;

QY 2 NEALMLSNGLILYLTVLFLEFLFSKVSVNLTKSK-KELTLFSTSNFLIMIAWTVMVNLF 60  
 Db 4 SEYMFULLSITIVYMT--KYLIFSLSDLTLPWKLQLTALALFF-----NQF 50

QY 61 ---YAEPLFYIASIYLNQRONSLNIFGPGVASSDFRR--AIFFLDGTG 112  
 Db 51 PYLSPLPLIDPFLFLVLUQKQLESKALKFLAVAPSVLVDLISRFMGTIVIPYFLSSG 110

QY 113 IVMSSI--ITTYMIEFAGIALSYFLSIVNDIGRLKDSITKMKVKRLLPNTMLY 170  
 Db 111 IVIGHIFDIALYLIFPSER-----INYMIGDKMTCQSGIKRSHNFQYQLMF 163

QY 171 YLIQIV-LVIVSYNVIPTKFR-----KEVVTYLFILFLTSFLSQYTKRQVN 220  
 Db 164 VLVVYDVFVING-TDPFLFHHSFLFVTPYKULMELLVLYLVSYFNHSSKYLN 222

QY 221 EIMAQKEAQRIITQSQSQTSLYKDRKSFRHDYLNITSLRIGIENKLASIEKYHQI 280  
 Db 223 ELRREQQAYMTLEIVGKHKLRYDVRAFQSDLSRERLSQAIKSESIQIQDYAQ 282

QY 281 LEKTHQHQLQDTNTIGHANLONDVKGTLASKTIEAQONKKAVNVVSSKIQLOPEMELL 340  
 Db 283 VHEANDYWDKHNKNISKLAKISAKSGIDLNVEPDNIKETYIPEL 342

QY 341 DFTTILSTLICDAIEAFESINPEQTLAFKNGSIVTIONSTKEQIDYSKIRENYS 400  
 Db 343 DULLIMSIFCDNAIEAARLEAQHOPHMSIAFLGLDQYMFVNTNTK KVVDINKIFEGYS 401

QY 401 TKSNSNRGIGAKVNHILEHYKTSQTSNHHFLPKROLL 438

Db 402 SKSBERGIGLSNAQRIKIRKPYLSLRTKSFDKERSQTL 439

RESULT 6

Q93YP4 PRELIMINARY; PRT; 451 AA.

AC DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 AC DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DE DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

GN SPY1605.

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus

OX NCBI\_TaxID=1314;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;

RX MEDLINE=2119264; PubMed=1129629;

RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
 PRIMEAUX C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
 Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J.,  
 Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.",  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).

DR EMBL; AE006491; AAC33321; -;

KW Kinase; Complete proteome;  
 SQ SEQUENCE 427 AA; 49908 MW; 01C9D236E4D7A0B6 CRC64;

Query Match 23.8%; Score 517; DB 16; Length 427;  
 Best Local Similarity 27.6%; Pred. No. 2.5e-22;  
 Matches 122; Conservative 110; Mismatches 184; Indels 26; Gaps 6;  
 Matches 122; Conservative 110; Mismatches 184; Indels 26; Gaps 6;

QY 5 LMILSNGLILYLTVLFLEFLFSKVSVNLTKSK-KELTLFSTSNFLIMIAWTVMVNLF 62  
 Db 1 MAVIYSMALIYATIFINSWIFAKVSAIKLSMKRVI-----IGISVIANMFDK 51



Db	236	EQ-RQLQLYDTEIVGLINEFRHDYAGMILSTQLTGNGSDMKEVERIFFHNLQSQANIS	294
Qy	288	LQPTRINIGHANTIONDAVKGILSKILEQNKKAVNNEVSSKIQLPEMLDFITLS	347
Db	295	IJKSDYVFFELNNDQIALSVLQKIFRAECGSVEVFMKDVTLEMKLJLFLVRAS	354
Qy	348	ILCDNATEAFAESLNPEIQLAFFFNGSTWITQNSTKQIDSKIFKENYSPKGSNRG	407
Db	355	VILNNNAVGEAABPSKMNNSVLSVLDKREIVFVQNSRQSYNILEEYEVGFSTKGENG	414
Qy	408	IGLAKVNHILHYPKSLQSNSHHHLFKOLLIK	441
Db	415	LGNNVKEITIDKYDEVLETDIETNYFIQVYRFK	448
RESULT	10		
P72446			
ID		PRELIMINARY;	PRT;
AC	P72446;		453 AA.
P72446;			
DT	01-FEB-1997	(TREMBLrel.	02, Created)
DT	01-FEB-1997	(TREMBLrel.	02, Last sequence update)
DT	01-DEC-2001	(TREMBLrel.	19, Last annotation update)
DE	HISTIDINE KINASE.		
GN	COM1.		
OS	Streptococcus gordonii.		
OC	Bacter; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC	Streptococcus.		
OX	NCBI_TaxID:1302;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPAIN-CHALLIS (NCTC 7868);		
RX	MEDLINE=97032151; PubMed=8878047;		
RA	Havarstein L.S., Gausel P., Nes I.F., Morrison D.A.;		
RT	"Identification of the streptococcal competence-phagomone receptor.";		
RL	Microbiol. 21:863-869(1995).		
DR	EMBL: X98109; CAA66787.1; -.		
KW	kinase.		
SQ	SEQUENCE	453 AA:	53633 MW: 281D55CDFB4063E2 CRC64;
Query Match	20.7%	Score	450; DB 2; Length 453;
Best Local Similarity	28.6%	Pred.	No. 1. 8e-18;
Matches	130;	Mismatches	180; Indels 50; Gaps
Qy	1	MNEALMLTISNGLTYLUVLFLFLFSKVSVNTLSKKELTIF-SISNLFNIMAVNVNL	59
Db	32	INIKLILFLWGLINVISFLYSLFL-----LLPDYWKFRDIFHYLEL-----	74
Qy	60	FYPAEPLPFIAISIYLQRNQLSNLNFYGL--LPVASSPDRRAIFFFDIGQIGVNG	116
Db	75	--IQPLIFYKFLIRKKEYENYLNLFLSFVYLSETSTELSVII-----SSTG	123
Qy	117	SSITT-YMIEPAGIASYFLSVEVDIGRLKDSLTKMY-KKLIPMNTMILYLYLT	174
Db	124	DSPVQHQHIDFTIINILSLFLILKVDFFDPYFEYIKEPIYKNDLWNVSYIVHIL	183
Qy	175	QVYI-----VYESV-NVPTPKFRKFVVYVYLFLILSFLSYTKDVKVONEMACKE	227
Db	184	NISHWFSERNAHNLSFASMAITFG-----IMFXSTFLYLSAREQEK --FLOQKE	235
Qy	228	AQPNITQYSQQLIESYKDRSRFHYNLNTLSRIGLAKLSTEKIHQOLEKTGQ	287
Db	236	EO-ROLOLTYDELVGLYNEIRGRHDYAGMLTSLOQINSGMDKEVERIFHNLVSQANS	294
Qy	288	LODTRYNIGHLANIQNDKVGILSAKILEAQNKKAVNVEVSSKQLPPEMELLPTILS	347
Db	295	LRSDTEFLFELVNQDPLARSLYVTFQKAFBECGVLFVEMKDVTFETPLPKLULVRAS	354
Qy	348	ILCONATEAFESLNPEIQLAFFFNGSTWITQNSTKQIDSKIFKENYSPKGSNRG	407
Db	355	VILNNNAVGEAABPSKMNNSVLSVLDKREIVFVQNSRQSYNILEEYEVGFSTKGENG	414
Qy	408	IGLAKVNHILHYPKSLQSNSHHHLFKOLLIK	441

DT	01-JAN-1998 (TREMBrel. 05, Last sequence update)	OC	Streptococcus
DT	01-DEC-2001 (TREMBrel. 19, Last annotation update)	OX	NCBL-TAXID=1303;
DE	HISTIDINE KINASE	RN	[1]
COMD		RP	SEQUENCE FROM N.A.
GN		RC	STRAIN-COL19;
OS	Streptococcus anginosus	RX	MEDLINE=9255539; PubMed=10322016;
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	RA	"Genetic diversity of the streptococcal competence (com) locus.";
OC	Streptococcus	RT	J. Bacteriol. 181:3144-3154(1999).
NCBI_TAXID	1328;	RL	[2]
RN		RN	SEQUENCE FROM N.A.
RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.
RC	SRAIN-NCTC 10713;	RC	STRAIN=COL19;
RX	MEDLINE=98012953; PubMed=9352304;	RX	MEDLINE=20098376; PubMed=10632854;
RA	Havarsen L.S., Hakenbeck R., Gaustad P.;	RA	Wittsell H.;
RT	"Natural competence in the genus Streptococcus: Evidence that streptococci can change phenotype by interspecies recombinational exchanges."	RT	"Chloroplast DNA variation and reticulate evolution in sexual and apomictic sections of dandelions.";
RR	Bacteriol. 179:6589-6594(1997).	RL	Mol. Ecol. 8:2023(2003)1999.
RL	EMBL: AU00564; CNA01342.1; -;	DR	EMBL: AJ240794; CAB39420.1; -;
DR	InterPro: IPR00594; HATPase_C	DR	InterPro: IPR04359; HIS_KIN_519.
DR	InterPro: IPR04359; HIS_KIN_519.	DR	SEQUENCE 441 AA; 51328 MN; 63AC03CAA3DEC9D CRC64;
KW	Pfam: PF02518; HATPase_c; 1.	DR	SEQUENCE 448 AA; 52922 MN; A995DB9B254ACB57 CRC64;
SQ	SEQUENCE 448 AA; 52922 MN; A995DB9B254ACB57 CRC64;	DR	SEQUENCE 441 AA; 51328 MN; 63AC03CAA3DEC9D CRC64;
Query Match	20.4%; Score 443.5; DB 2; Length 448; Best Local Similarity 23.6%; Pred: No. 4.3e-18; Matches 105; Conservative 113; Mismatches 195; Indels 31; Gaps 8;	Query Match	20.2%; Score 439.5; DB 2; Length 441; Best Local Similarity 27.6%; Pred: No. 7.1e-18; Matches 121; Conservative 89; Mismatches 171; Indels 57; Gaps 11;
QY	12 LITYLTYLLEFLFSKVNMSNLTKELTLISSLNMLIMAVMVNLVEP--AEPLYF 68	QY	33 LSKEETLFLSSNLIMAVMVNLVEP-----PL---YFIALSIVLNRONS
Db	10 IFSITEAVSVVYCYKKISRVNRKVNTIHTLCLGTVFSTDPTLHYSTRYMMFPQLPF 69	Db	28 IKSKELYFFGTYITLIVE-AVLELSFYLYLDELIVIERFLPFLGLYSYFRNIKQY-ERDKG
QY	69 LAISIVLNQNQNLISNLNFYGL--IYVASSDLFRAAIIFILDCTQGIVMGSSIIYMI 125	QY	81 USLNIFYGLPVASSDLFRAAIIFILDCTQGIVMGSSIIYMI-----TIMIEF 127
Db	70 Y-LYFFKVKVHLFLAFALSLVLAASSTEFVIT-----SSYVGDFKVDRW- 119	Db	86 LFSLSLILSYESTH-----FLSVTSITSGDNFVSQYHDPFFFTVLTYFWL 136
QY	126 EFAGIALSYLFSLVFN-----VDIGRLK-DSLTKMKVKRKLIPMNTMLYLQLQVL 177	QY	128 AGIALSYLFSLVFNVDIGRLKDSLTKMKVKRKLIPMNTMLYLQLQVIVIESVNP 187
Db	120 ----GLFYIFTNITISLFLVFLKADYFKENKFYKFKRIDEKFRENTIVOLNFVLFTHLLNS 174	Db	137 KLIBPHLELSYFDY-----LYPELKVKVFALLLHIVSFV-----DMVS
QY	178 YVIESYNVNPYTKFRKEVIVIYLILILISLFLSFLSQYTQKVQNEITMAQEAOIRNTQ 237	QY	188 TLK-FRKFVVIVYLTF--LILTSFLSFLSQYTQKVQNEITMAQEAOIRNTQYSQIESL 243
Db	175 HWLIS--NNKLNLSFSSMIAITCFLMFMSILYLOQSIREKEPOIKREREQLOORYT 232	Db	180 TIKHUNSEFGSILSIVFLSILLTPEFAMKSHBQMEKATAKQKBFQKHQNYTDIEVL 239
QY	238 QTESLYKDIDRFRHDYLNILTSLRIGIENKLASIEKIHQIJEKTGHOLQDTRYNIGH 297	QY	244 YKDTSFRHYNLTSLRIGIENKLASIEKIHQIJEKTGHOLQDTRYNIGHANION 303
Db	233 DELTSLVNEIRPFRHDYGGMLASFQSAIHGDIKEVERTYQEVLYNQNLRSKDYTFD 292	Db	240 YNEIRGFRHDYAGMVSOMAIDSGDLOEIDRYNEVLUKANKLRSKDYTFDINTED 299
QY	298 LANITONDVAKGILSAKILEAQONKKIAYNEVSSKIQOLPEMELDFTTILSICDNAEAKTESLNP 357	QY	304 DAVKGITLSAKILEAQONKKIAYNEVSSKIQOLPEMELDFTTILSICDNAEAKTESLNP 363
Db	293 LNNVGDALSARWMTQTEKTRYNIELTEFQKVNPPLPIKLDLVRMVTSULLNNTGA 352	Db	300 SALRSLVQASIVWNRNGCFTEVKODITKIELLDLVRMVSLANNVEGSDSYK 359
QY	358 FESLNPETQLOAFFRKNGNSIVFTIONSTREKOIDVSKERIFKENYSTKGSNRGIGAKVNHT 417	QY	364 EIQLAFKKGSIVFTIONSTREKOIDVSKERIFKENYSTKGSNRGIGAKVNHTLHYPK 423
Db	353 AESYQKTMVNLVLDPEFTIVLQNSKRRFLDLEIYQDFSTKGEGRGLGSLNIKEI 412	Db	360 OMEVAVKMETEVIVIONSKWMTPSGDIFALGFSTKGRNRGVGLNNVKEELDKNYI 419
OY	418 EHYPKTSQTSNHHHLFQLLTIK 441	QY	424 SLQTSHHHHFQLLTIK 441
Db	413 NYVGGIILDTKIEDEYFQVMVR 436	Db	420 ILETMEGSTFROJIRFK 437
RESULT	13	RESULT	14
Q9X95	PRELIMINARY;	Q9X90	PRELIMINARY;
ID	Q9X95	AC	Q9X90
AC	Q9X95;	DT	01-NOV-1999 (TREMBrel. 12, Created)
DT	01-NOV-1999 (TREMBrel. 12, Last sequence update)	DT	01-NOV-1999 (TREMBrel. 12, Last sequence update)
DT	01-Nov-1999 (TREMBrel. 12, Last sequence update)	DT	01-DEC-2001 (TREMBrel. 19, Last annotation update)
DE	COMD PROTEIN.	DE	COMD PROTEIN.
GN		GN	
OS	Streptococcus mitis.	OS	Streptococcus mitis.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC	Streptococcus	OC	Streptococcus
OX	NCBI_TAXID=28037;	OX	NCBI_TAXID=28037;
RN		RN	

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

